(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 December 2002 (27.12.2002)

PCT

(10) International Publication Number
WO 02/102323 A2

(51) International Patent Classification[7]: A61K

(21) International Application Number: PCT/US02/19560

(22) International Filing Date: 14 June 2002 (14.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/298,296 14 June 2001 (14.06.2001) US

(71) Applicant *(for all designated States except US)*: **BRISTOL-MYERS SQUIBB COMPANY** [US/US]; P.O BOX 4000, ROUTE 206 and PROVINCELINE ROAD, PRINCETON, NJ 08543-4000 (US).

(72) **Inventors; and**

(75) **Inventors/Applicants** *(for US only)*: **JACKSON, Donald, G.** [US/US]; 2641 Main St. Apt. 1, Lawrenceville, NJ 08648 (US). **LORENZI, Matthew, V.** [US/US]; 424 South 7th Street, Philadelphia, PA 19147 (US). **ATTAR, Ricardo, M.** [US/US]; 10 Santina Ct., Lawrenceville, NJ 08648 (US). **GOTTARDIS, Marco** [US/US]; 9 Harris Road, Princeton, NJ 08540 (US).

(74) **Agents: D'AMICO, Stephen** et al.; Bristol-Myers Squibb Company, P.O. Box 4000, Route 206 and Provinceline Road, Princeton, NJ 08543-4000 (US).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) **Title:** NOVEL HUMAN HISTONE DEACETYLASES

```
    GlyIleAlaTyrAspProLeuMetLeuLysHisGlnCysValCysGly
1   ggaattgcctatgaccccttgatgctgaaacaccagtgcgtttgtggc
    ccttaacggatactggggaactacgactttgtggtcacgcaaacaccg

    AsnSerThrThrHisProGluHisAlaGlyArgIleGlnSerIleTrp
49  aattccaccacccaccctgagcatgctggacgaatacagagtatctgg
    ttaaggtggtgggtgggactcgtacgacctgcttatgtctcatagacc

    SerArgLeuGlnGluThrGlyLeuLeuAsnLysCysGluArgIleGln
97  tcacgactgcaagaaactgggctgctaaataaatgtgagcgaattcaa
    agtgctgacgttctttgacccgacgatttatttacactcgcttaagtt

    GlyArgLysAlaSerLeuGluGluIleGlnLeuValHisSerGluHis
145 ggtcgaaaagccagcctggaggaaatacagcttgttcattctgaacat
    ccagcttttcggtcggacctcctttatgtcgaacaagtaagacttgta

    HisSerLeuLeuTyrGlyThrAsnProLeuAspGlyGlnLysLeuAsp
193 cactcactgttgtatggcaccaacccccctggacggacagaagctggac
    gtgagtgacaacataccgtggttggggggacctgcctgtcttcgacctg

    ProArgIleLeuLeuGlyAspAspSerGlnLysPhePheSerSerLeu
241 cccaggatactcctaggtgatgactctcaaaagtttttttcctcatta
    gggtcctatgaggatccactactgagagttttcaaaaaaaggagtaat

    ProCysGlyGlyLeuGlyValSerThr
289 ccttgtggtggacttggggtaagtaca
    ggaacaccacctgaaccccattcatgt
```

(57) **Abstract:** The present invention relates to newly discovered human histone deacetylases (HDACs), also referred to as histone deacetylase-like polypeptides. The polynucleotide sequences and encoded polypeptides of the novel HDACs are encompassed by the invention, as well as vectors comprising these polynucleotides and host cells comprising these vectors. The invention also relates to antibodies that bind to the disclosed HDAC polypeptides, and methods employing these antibodies. Also related are methods of screening for modulators, such as inhibitors or antagonists, or agonists. The invention also relates to diagnostic and therapeutic applications which employ the disclosed HDAC polynucleotides, polypeptides, and antibodies, and HDAC modulators. Such applications can be used with diseases and disorders associated with abnormal cell growth or proliferation, cell differentiation, and cell survival, e.g., neoplastic cell growth, and especially breast and prostate cancers or tumors.

WO 02/102323 A2

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

NOVEL HUMAN HISTONE DEACETYLASES

RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Application Serial No. 60/298,296, filed June 14, 2001, which is incorporated by reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to novel members of the histone deacetylase (HDAC) family, including BMY_HDAL1, BMY_HDAL2, BMY_HDAL3, BMY_HDACX_v1, BMY_HDACX_v2, and HDAC9c. Specifically related are nucleic acids encoding the polypeptide sequences, vectors comprising the nucleic acid sequences, and antibodies that bind to the encoded polypeptides. In addition, the invention relates to pharmaceutical compositions and diagnostic reagents comprising one or more of the disclosed HDAC components. The present invention also relates to methods of treating a disease or disorder caused by malfunction of an HDAC, e.g., due to mutation or altered gene expression. The invention further relates to methods of using a modulator of an HDAC of the present invention to treat or ameliorate a disease state. Also related are methods for devising antisense therapies and prophylactic treatments using the HDACs of the invention. In particular, the disclosed HDAC components and methods may be used to prevent, diagnose, and treat diseases and disorders associated with abnormal cell growth or proliferation, cell differentiation, or cell survival, e.g., neoplasias, cancers, and tumors, such as breast and prostate cancers or tumors, and neurodegerative diseases.

BACKGROUND OF THE INVENTION

Chromatin is a dynamic protein-DNA complex which is modulated by post-translational modifications. These modifications, in turn, regulate cellular processes such as gene transcription and replication. Key chromatin modifications include the acetylation and deacetylation of nucelosomal histone proteins. Acetylation is catalyzed by histone acetylases (HATs), whereas deacetylation is catalyzed by deacetylases (HDACs or HDAs). HDACs catalyze the removal of acetyl groups from the N-termini of histone

core proteins to produce more negatively charged chromatin. This results in chromatin compaction, which shuts down gene transcription. In addition, inhibition of HDACs results in the accumulation of hyperacetylated histones. This, in turn, is implicated in a variety of cellular responses, including altered gene expression, cell differentiation, and cell-cycle arrest (see, generally, S.G. Gray et al., 2001, *Exp. Cell Res.* 262(2):75-83, and U.S. Patent Nos. 6,110,697 and 6,068,987 to Dulski et al.).

The HDAC gene family is composed of two distinct classes. Class I HDACs are related to the yeast transcriptional regulator, RPD3. Class II HDACs include a subgroup of proteins containing a C-terminal catalytic domain as well as a separate N-terminal domain with transcriptional repression activity. Class III HDAC proteins are related to the yeast sir2 protein and require NAD for activity. Class I HDACs are predominantly nuclear, whereas class II HDACs are transported between the cytoplasm and nucleus as part of the regulation of cellular proliferation and/or differentiation (reviewed in S. Khochbin et al., 2001, *Curr. Opin. Genet. Dev.* 11(2):162-6).

The best characterized substrates for HDACs include histone or histone-like peptide sequences containing N-terminal lysines. However, non-histone HDAC substrates have also been identified, including several transcription factors. Non-histone substrates for HDACs include p53, androgen receptor, LEF1/TCF4 (B.R. Henderson et al., 2002, *J. Biol. Chem.*, published online on May 1, 2002 as Manuscript M110602200), GATA-1, and estrogen receptor-alpha (reviewed in D.M. Vigushin et al., 2002, *Anticancer Drugs* 13(1):1-13). For these substrates, deacetylation has been shown to regulate DNA/protein interactions or protein stability. Such molecules may therefore represent therapeutic targets of HDACs. Importantly, the histone deacetylase function of HDACs represses transcription by removing the acetyl moieties from amino terminal lysines on histones, thereby resulting in a compact chromatin structure. In contrast, the non-histone deacetylase function of HDACs can either repress or activate transcription.

There has been considerable interest in modulating the activity of HDACs for the treatment of a variety of diseases, particularly cancer. Several

small molecule inhibitors of HDAC have shown anti-proliferative activities on a number of tumor cell lines and potent anti-tumor activity in pre-clinical tumor xenograft models, most recently, CBHA (D.C. Coffey et al., 2001, *Cancer Res.* 61(9):3591-4), pyroxamide, (L.M. Butler et al, 2001, *Clin. Cancer Res.* 7(4):962-70), and CHAP31 (Y. Komatsu et al., 2001, *Cancer Res.* 61(11):4459-66). Several inhibitors are presently being evaluated as single agents and in combination regimens with cytotoxic agents for the treatment of advanced malignancies (reviewed in P.A. Marks et al., *Curr. Opin. Oncol.* 2001 Nov;13(6):477-83). Thus, HDAC inhibitors are being developed as anti-tumor agents, as well as agents useful for gene therapy (McInerney et al., 2000, *Gene Ther.* 7(8):653-663).

Small molecule inhibitors of HDAC activity that have undergone extensive analysis include trichostatin A (TSA), trapoxin, SAHA (V.M. Richon et al., 2001, *Blood Cells Mol. Dis.* 27(1):260-4), CHAPs (Y. Komatsu et al., 2001, *Cancer Res.* 61(11):4459-66), MS-27-275 (reviewed in M. Yoshida et al., 2001, *Cancer Chemother. Pharmacol.* 48 Suppl. 1:S20-6), depsipeptide (FR901228; FK228; see, e.g., V. Sandor et al., 2002, *Clin. Cancer Res.* 8(3):718-28), and CI-994 (see, e.g., P.M. LoRusso et al., 1996, *New Drugs* 14(4):349-56; S. Prakash et al., 2001, *Invest. New Drugs* 19(1):1-11). Trichostatin A and trapoxin have been reported to be reversible and irreversible inhibitors, respectively, of mammalian histone deacetylase (Yoshida et al, 1995, *Bioassays*, 17(5):423-430). Trichostatin A has also been reported to inhibit partially purified yeast histone deacetylase (Sanchez del Pino et al., 1994, *Biochem. J.*, 303:723-729). Moreover, trichostatin A is an antifungal antibiotic and has been shown to have anti-trichomonal activity and cell differentiating activity in murine erythroleukemia cells, as well as the ability to induce phenotypic reversion in ras-transformed fibroblast cells (see e.g. U.S. Pat. No. 4,218,478; and Yoshida et al., 1995, *Bioassays*, 17(5):423-430, and references cited therein). Trapoxin A, a cyclic tetrapeptide, induces morphological reversion of v-sis-transformed NIH/3T3 cells (Yoshida and Sugita, 1992, *Jap. J. Cancer Res.*, 83(4):324-328).

The therapeutic effects of HDAC inhibition are believed to occur through the induction of differentiation and/or apoptosis through the up-regulation of genes such as the cyclin dependent kinase inhibitors, p21 and p27 (see, e.g., W. Wharton et al., 2000, *J. Biol. Chem.* 275(43):33981-7; L. Huang et al., 2000, *Mol. Med.* 6(10):849-66). Although known HDAC inhibitors are efficacious as anti-tumor agents, they are also associated with toxicity (see, e.g., V. Sandor et al., 2002, *Clin. Cancer Res.* 8(3):718-28). Such toxicity is believed to be caused by a non-selective mechanism of targeting multiple HDACs. Despite the potent anti-tumor activity of HDAC inhibitors, it is still unclear which HDACs are necessary to produce an anti-proliferative response. Furthermore, little progress has been made in comparing the HDAC gene expression profiles in tumor versus normal cells. Differential HDAC expression may underlie the tumor-selective responses of HDAC inhibition. In addition, a cellular growth advantage may be conferred by the expression of particular HDACs. Therefore, there is a need for further insight into the consequences of selective HDAC inhibition, or activation.

SUMMARY OF THE INVENTION

The present invention provides novel histone deacetylase (HDAC) nucleic acid sequences and their encoded polypeptide products, also called histone deacetylase like (HDAL) sequences and products herein, as well as methods and reagents for modulating HDACs.

It is an aspect of this invention to provide new HDAC nucleic acid or protein sequences, or cell lines overexpressing HDAC nucleic acid and/or encoded protein, for use in assays to identify small molecules which modulate HDAC activity, preferably antagonize HDAC activity.

It is another aspect of the present invention to employ HDAC protein structural data for the *in silico* identification of small molecules which modulate HDAC activity. This structural data could be generated by experimental techniques (for example, X-Ray crystallography or NMR spectroscopy) or by computational modeling based on available histone deacetylase structures (for example, M.S. Finnin et al., 1999, *Nature*, 401(6749):188-193).

Another aspect of the present invention provides modulators of HDAC activity, e.g., antagonists or inhibitors, and their use to treat neoplastic cells, e.g., cancer cells and tumor cells. In one aspect of the invention, breast or prostate cancers or tumors are treated using the HDAC modulators. The modulators of the invention can be employed alone or in combination with standard anti-cancer regimens for neoplastic cell, e.g., tumor and cancer, treatments.

In addition, the present invention provides diagnostic reagents (i.e., biomarkers) for the detection of cancers, tumors, or neoplastic growth. In one embodiment, HDAC (e.g., HDAC9c) nucleic acids or anti-HDAC antibodies are used to detect the presence of specific cancers or tumors, such as breast or prostate cancers or tumors.

It is yet another aspect of the present invention to employ HDAC inhibitors in the regulation of the differentiation state of normal cells such as hematopoietic stem cells. According to this invention, a method is provided for the use of modulators of HDAC in *ex vivo* therapies, particularly as a means to modulate the expression of gene therapeutic vectors.

Yet another aspect of this invention is to provide antisense nucleic acids and oligonucleotides for use in the regulation of HDAC and HDAL gene transcription or translation.

An additional aspect of this invention pertains to the use of HDAC nucleic acid sequences and antibodies directed against the produced protein for prognosis or susceptibility for certain disorders (e.g., breast or prostate cancer).

Further aspects, features and advantages of the present invention will be better appreciated upon a reading of the detailed description of the invention when considered in connection with the accompanying figures/drawings.

BRIEF DESCRIPTION OF THE FIGURES

The file of this patent contains at least one figure executed in color. Copies of this patent with color figure(s) will be provided by the Patent and Trademark Office upon request and payment of the necessary fee.

FIG. 1 shows the novel BMY_HDAL1 partial nucleic acid (cDNA) sequence (SEQ ID NO:1) and the encoded amino acid sequence (SEQ ID NO:2) of the BMY_HDAL1 polypeptide product. The top line in each group of Fig. 1 presents the BMY_HDAL1 protein sequence (SEQ ID NO:2) in 3-letter IUPAC form; the middle line presents the nucleotide sequence of the BMY_HDAL1 coding strand (i.e., SEQ ID NO:1); and the bottom line presents the nucleotide sequence of the reverse strand (SEQ ID NO:3).

FIGS. 2A and 2B show the amino acid sequences of the novel histone deacetylase-like proteins BMY_HDAL1 (SEQ ID NO:2), BMY_HDAL2 (SEQ ID NO:4) and BMY_HDAL3 (SEQ ID NO:5) aligned with the following known histone deacetylase proteins: *S. cerevisiae* HDA1 (SC_HDA1), (SEQ ID NO:6); human HDAC4 (HDA4), (SEQ ID NO:7); human HDAC5 (HDA5), (SEQ ID NO:8); human HDAC7 (HDA7), (SEQ ID NO:9) and to a histone deacetylase-like protein ACUC from *Aquifex aeolicus* (AQUIFEX_HDAL), (SEQ ID NO:10), (M.S. Finnin et al., 1999, *Nature*, 401(6749):188-193). Residues identical among all proteins are in shown in black text on a gray background. The sequences were aligned using the ClustalW algorithm as implemented in the VectorNTI sequence analysis package (1998, 5.5 Ed., Informax, Inc.) with a gap opening penalty of 10, a gap extension penalty of 0.1 and no end gap penalties.

FIGS. 3A and 3B show a GenewiseDB comparison of BMY_HDAL1 amino acid sequence (SEQ ID NO:2) and human HDAC5 (HDA5) amino acid sequence (SEQ ID NO:8). Genewise results from HDA5_HUMAN_run2 applied to AC002088 nucleic acid (coding) sequence. (SEQ ID NO:11).

FIG. 4 presents the results of sequence motif analysis of motifs within the BMY_HDAL1 amino acid sequence.

FIG. 5 shows the novel BMY_HDAL2 partial nucleic acid (cDNA) sequence (SEQ ID NO:12) and the encoded amino acid sequence (SEQ ID NO:4) of the BMY_HDAL2 polypeptide product. The top line in each group of Fig. 5 presents the BMY_HDAL2 protein sequence (SEQ ID NO:4) in 3-letter IUPAC form; the middle line presents the nucleotide sequence of the

BMY_HDAL2 coding strand (i.e., SEQ ID NO:12); and the bottom line presents the nucleotide sequence of the reverse strand (SEQ ID NO:13).

FIG. 6 presents a GenewiseDB comparison of the BMY_HDAL2 amino acid sequence (SEQ ID NO:4) and human HDAC5 (HDA5) amino acid sequence (SEQ ID NO:8). Genewise results from HDA5_HUMAN_run3 applied to AC002410 nucleic acid sequence (SEQ ID NO:14).

FIG. 7 shows PROSITE motifs identified in the predicted amino acid sequence of the novel BMY_HDAL2 (SEQ ID NO:4). MOTIFS are from: bmy_hdal2.aa.fasta.

FIGS. 8A and 8B show the sequences of the N- and C-terminal sequences of BMY_HDAL3 as determined from BAC AC004994 and BAC AC004744. **FIG. 8A** presents the most N-terminal region of the BMY_HDAL3 amino acid sequence (SEQ ID NO:15) presented herein as encoded by the human genomic BAC AC004994 polynucleotide sequence (SEQ ID NO:17). **FIG. 8B** presents an additional C-terminal portion of the BMY_HDAL3 amino acid sequence (SEQ ID NO:16) as encoded by human genomic BAC AC004744 polynucleotide sequence (SEQ ID NO:18).

FIG. 9 shows partial transcripts identified from the AC004994 polynucleotide sequence (SEQ ID NO:17) and from the AC004744 polynucleotide sequence (SEQ ID NO:18) assembled into a single contig, which was designated BMY_HDAL3 (SEQ ID NO:19) using the VectorNTI ContigExpress program (Informax, Inc.).

FIG. 10 presents the BMY_HDAL3 partial nucleic acid sequence (SEQ ID NO:19) and the encoded amino acid sequence (SEQ ID NO:5) based on the assembled BMY_HDAL3 sequence described in FIG. 9. The top line in each group of FIG. 10 presents the BMY_HDAL3 protein sequence (SEQ ID NO:5) in 3-letter IUPAC form; the middle line presents the nucleotide sequence of the BMY_HDAL3 coding strand (i.e., SEQ ID NO:19); and the bottom line presents the nucleotide sequence of the reverse strand (SEQ ID NO:20).

FIG. 11 presents the results of the GCG Motifs program used to analyze the BMY_HDAL3 partial predicted amino acid sequence for motifs in

the PROSITE collection (K. Hofmann et al., 1999, *Nucleic Acids Res.*, 27(1):215-219) with no allowed mismatches.

FIG. 12 shows a multiple sequence alignment of the novel human HDAC, BMY_HDAL3, amino acid sequence (SEQ ID NO:5) with the amino acid sequence of AAC78618 (SEQ ID NO:21) and with the amino acid sequence of AAD15364 (SEQ ID NO:22). AAC78618 is a histone deacetylase-like protein predicted by genefinding and conceptual translation of AC004994 and which was entered in Genbank. AAD15364 is a similar predicted protein derived from AC004744 and entered in Genbank. AAC78618, AAD15364 and BMY_HDAL3 were aligned using the ClustalW algorithm as implemented in the VectorNTI sequence analysis package (1998, 5.5 Ed., Informax, Inc.) with a gap opening penalty of 10, a gap extension penalty of 0.1 and no end gap penalties. Residues identical among all proteins are shown in white text on a black background; conserved residues are shown in black text on a gray background.

FIG. 13 shows a BLASTN alignment of the AA287983 polynucleotide sequence (SEQ ID NO:23) and BMY_HDAL3 polynucleotide sequence from SEQ ID NO:19. Genbank accession AA287983 is a human EST sequence (GI # 1933807; Incyte template 1080282.1) which was identified by BLASTN searches against the Incyte LifeSeq database using the NCBI Blast algorithm (S.F. Altschul et al., 1997, *Nucl. Acids Res.*, 25(17):3389-3402) with default parameters. The AA287983 human EST was isolated from a germinal B-cell library. No additional ESTs are included in the Incyte template derived from this cluster (Incyte gene ID 180282).

FIGS. 14A-14H present other histone deacetylase sequences, as shown in FIGS. 2A and 2B. **FIG. 14A**: *Aquifex* ACUC protein amino acid sequence (SEQ ID NO:10); **FIG. 14B**: *Saccharomyces cerevisiae* histone deacetylase 1 amino acid sequence (SEQ ID NO:6); **FIG. 14C**: *Homo sapiens* histone deacetylase 4 amino acid sequence (SEQ ID NO:7); **FIG. 14D**: *Homo sapiens* histone deacetylase 5 amino acid sequence (SEQ ID NO:8); **FIG. 14E**: *Homo sapiens* histone deacetylase 7 amino acid sequence (SEQ ID NO:9); **FIG. 14F**: Human EST AA287983 nucleic acid sequence

(SEQ ID NO:23); **FIG. 14G**: Human predicted protein AAD15364 amino acid sequence(SEQ ID NO:22); and **FIG. 14H**: Human predicted protein AAC78618 amino acid sequence (SEQ ID NO:21).

FIGS. 15A-15C depict the nucleotide and amino acid sequence information for HDAC9c. The polypeptide sequence (SEQ ID NO:87) is shown using the standard 3-letter abbreviation for amino acids. The DNA sequence (SEQ ID NO:88) of the coding strand is also shown. **FIGS. 15D-15F** depict an amino acid sequence alignment of HDAC9c. The predicted amino acid sequence of HDAC9c (SEQ ID NO:87) was aligned to previously identified HDACs, including HDAC9 (AY032737; SEQ ID NO:89), HDAC9a (AY032738; SEQ ID NO:90), and HDAC4 (ALF132608; SEQ ID NO:91), using ClustalW (D.G. Higgins et al., 1996, *Methods Enzymol.* 266:383-402). Identical amino acids are shown in white text on a black background; conserved amino acids are shown in black text on a gray background.

FIGS. 16A-16C depict expression levels of HDAC9 in human cancer cell lines and normal adult tissue. **FIG 16A**: Northern blot analysis of HDAC9 expression in normal adult tissue. **FIG 16B**: Quantitative PCR mRNA analysis of HDAC9 expression in human tumor cell lines. **FIG 16C**: Nuclease protection assay analysis of HDAC9 expression in human tumor cell lines. **FIG. 16D** shows the nucleotide sequence of HDAC9c used to derive the probes used for Northern blotting and nuclease protection analysis (SEQ ID NO:92). The probes were derived from the HDAC9c nucleotide sequence, and were predicted to hybridize to HDAC9c and HDAC9 (AYO32737), but not HDAC9a (AYO32738).

FIGS. 17A-17C illustrate the increase of HDAC9 gene expression in human cancer tissues. **FIGS. 17A-17B**: Summary of HDAC9 expression in selected tissues, as assayed by *in situ* hybridization. **FIG. 17C**: Photomicrographs of representative cells showing HDAC9 or actin staining.

FIG. 18 shows HDAC9c-mediated induction of morphological transformation of NIH/3T3 cells. The panels show photomicrographs of soft agar growth of vector (upper panel), FGF8 (middle panel) and HDAC9c (lower panel) transfected NIH/3T3 cells. Cells are shown at 10 X magnification.

FIG. 19 shows HDAC9c induction of actin stress fiber formation in NIH/3T3 cells. Stable NIH/3T3 cells expressing the indicated constructs were stained with phalloidin-TRITC and visualized by fluorescent microscopy.

FIGS. 20A-20C depict the nucleotide and amino acid sequence information for BMY_HDACX variant 1, also called BMY_HDACX_v1 and HDACX_v1. BMY_HDACX_v1 represents a partial cDNA sequence obtained from cells expressing a transcript variant of human HDAC9. The polypeptide sequence (SEQ ID NO:93) is shown using the standard 3-letter abbreviation for amino acids. The DNA sequence (SEQ ID NO:94) of the coding strand is also shown.

FIGS. 21A-21B depict the nucleotide and amino acid sequence information for BMY_HDACX variant 2, also called BMY_HDACX_v2 and HDACX_v2. BMY_HDACX_v2 represents a full-length sequence of a novel transcript variant (i.e., splice product) of HDAC9. The polypeptide sequence (SEQ ID NO:95) is shown using the standard 3-letter abbreviation for amino acids. The DNA sequence (SEQ ID NO:96) of the coding strand is also shown.

FIGS. 22A-22I depict the nucleotide and amino acid sequence information for the previously identified HDAC9 transcript variants. **FIGS. 22A-22C**: HDAC9 variant 1 (HDAC9v1; NCBI Ref. Seq. NM_058176). The polypeptide sequence (SEQ ID NO:89) is shown using the standard 3-letter abbreviation for amino acids. The DNA sequence (SEQ ID NO:97) of the coding strand is also shown. **FIGS. 22D-22F**: HDAC9 variant 2 (HDAC9v2; NCBI Ref. Seq. NM_058177). The polypeptide sequence (SEQ ID NO:90) is shown using the standard 3-letter abbreviation for amino acids. The DNA sequence (SEQ ID NO:98) of the coding strand is also shown. **FIGS. 22G-22I**: HDAC9 variant 3 (HDAC9v3; NCBI Ref. Seq. NM_014707). The polypeptide sequence (SEQ ID NO:99) is shown using the standard 3-letter abbreviation for amino acids. The DNA sequence (SEQ ID NO:100) of the coding strand is also shown.

FIGS. 23A-23K depict a multiple sequence alignment of nucleotide sequences representing known and novel HDAC9 splice products. The

cDNAs for BMY_HDACX_v1 (SEQ ID NO:94) and BMY_HDACX_v2 (SEQ ID NO:96) nucleotide sequences were aligned to the three reported splice products of the HDAC9 gene, including HDAC9v1 (NCBI Ref. Seq. NM_058176; SEQ ID NO:97), HDAC9v2 (NCBI Ref .Seq. NM_058177; SEQ ID NO:98), and HDAC9v3 (NCBI Ref. Seq. NM_014707; SEQ ID NO:100) using the sequence alignment program ClustalW (D.G. Higgins et al., 1996, *Methods Enzymol.* 266:383-402). The consensus sequence is shown on the bottom line (SEQ ID NO:106). Identical nucleotides are shown in white text on a black background. Selected splice junctions are indicated below the alignment; these junctions were identified by comparison of the cDNA sequences to the assembled genomic contig NT_00798.1 using the Sim4 algorithm (L. Florea et al., 1998, *Genome Res.* 8:967-74). It is noted that the HDAC9 (AY032737) nucleotide and amino acid sequences are identical to the HDAC9v1 (NM_058176) nucleotide and amino acid sequences. Similarly, the HDAC9a (AY032738) nucleotide and amino acid sequences are identical to the HDAC9v2 (NM_058177) nucleotide and amino acid sequences.

FIGS. 24A-24D depict a multiple sequence alignment of amino acid sequences representing known and novel HDAC polypeptides. The amino acid sequences encoded by transcript variants BMY_HDACX_v1 (SEQ ID NO:93) and BMY_HDACX_v2 (SEQ ID NO:95) were aligned to amino acid sequences encoded by known splice variants of human histone deacetylase 9 including HDAC9v1 (NCBI Ref. Seq. NM_058176; SEQ ID NO:89), HDAC9v2 (NCBI Ref .Seq. NM_058177; SEQ ID NO:90), and HDAC9v3 (NCBI Ref. Seq. NM_014707; SEQ ID NO:99), and to human histone deacetylases 4 and 5 (HDA5, SEQ ID NO:8; HDA4, SEQ ID NO:7) using the multiple sequence alignment program ClustalW (D.G. Higgins et al., 1996, *Methods Enzymol.* 266:383-402). The consensus sequence is shown on the bottom line (SEQ ID NO:107). Residues conserved among all polypeptides are shown in white text on a black background; residues conserved in a majority of polypeptides are shown in black text on a gray background.

FIGS. 25A-25C depict a multiple sequence alignment of amino acid sequences showing novel HDAC polypeptides. The amino acid sequences of

BMY_HDAL1 (SEQ ID NO:2), BMY_HDAL2 (SEQ ID NO:4), BMY_HDAL3 (SEQ ID NO:5), HDAC9c (SEQ ID NO:87), HDACX_v1 (SEQ ID NO:93), and HDACX_v2 (SEQ ID NO:95) were aligned using the T-Coffee program (C. Notredame et al., 2000, *J. Mol. Biol.* 302:205-217; C. Notredame et al., 1998, *Bioinformatics* 14:407-422). Identical residues are shown in black text on a gray background.

DESCRIPTION OF THE INVENTION

The present invention discloses several novel HDAC nucleotide sequences and encoded products. New members of the histone deacetylase protein family have been identified as having identity to known HDACs. Three new HDACs are referred to as BMY_HDAL1, BMY_HDAL2, and BMY_HDAL3 herein, wherein HDAL signifies histone deacetylase like proteins in current nomenclature. These proteins are most similar to the known human histone deacetylase, HDAC9. Novel HDAC9 splice variants, termed HDACX_v1 and HDACX_v2, have also been identified. In addition, HDAC9c, an HDAC9-related family member, has been newly identified and cloned. The nucleic acid sequences encoding the novel HDAC polypeptides are provided together with the description of the means employed to obtain these novel molecules. Such HDAC products can serve as protein deacetylases, which are useful for disease treatment and/or diagnosis of diseases and disorders associated with cell growth or proliferation, cell differentiation, and cell survival, e.g., neoplastic cell growth, cancers, and tumors.

As shown herein, HDAC9 expression is elevated in tumor cell lines, as determined by quantitative PCR analysis. Elevated expression of HDAC9 was also observed in clinical specimens of human tumor tissue compared to normal tissue, using *in situ* hybridization (ISH) and an HDAC9-specific riboprobe. Further, cell biological assessment of HDAC9c revealed that overexpression of HDAC9c confers a growth advantage to normal fibroblasts. These results indicate that HDAC9c can be used as a diagnostic marker for tumor progression and that selective HDAC9c inhibitors can be used to target specific cancer or tumor types, such as breast and prostate cancers or tumors.

Definitions

The following definitions are provided to more fully describe the present invention in its various aspects. The definitions are intended to be useful for guidance and elucidation, and are not intended to limit the disclosed invention and its embodiments.

HDAC polypeptides (or proteins) refer to the amino acid sequence of isolated, and preferably substantially purified, human histone deacetylase proteins isolated as described herein. HDACs may also be obtained from any species, preferably mammalian, including mouse, rat, non-human primates, and more preferably, human; and from a variety of sources, including natural, synthetic, semi-synthetic, or recombinant. The probes and oligos described may be used in obtaining HDACs from mammals other than humans. The present invention more particularly provides six new human HDAC family members, namely, BMY_HDAL1, BMY_HDAL2, BMY_HDAL3, HDACX_v1, HDACX_v2, and HDAC9c, their polynucleotide sequences (e.g., SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, SEQ ID NO:96, and sequences complementary thereto), and encoded products (e.g., SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, and SEQ ID NO:95).

An agonist (e.g., activator) refers to a molecule which, when bound to, or interactive with, an HDAC polypeptide, or a functional fragment thereof, increases or prolongs the duration of the effect of the HDAC polypeptide. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules that bind to and modulate the effect of an HDAC polypeptide. An antagonist (e.g., inhibitor, blocker) refers to a molecule which, when bound to, or interactive with, an HDAC polypeptide, or a functional fragment thereof, decreases or eliminates the amount or duration of the biological or immunological activity of the HDAC polypeptide. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules that decrease, reduce or eliminate the effect and/or function of an HDAC polypeptide.

"Nucleic acid sequence", as used herein, refers to an oligonucleotide, nucleotide, or polynucleotide (e.g., DNA, cDNA, RNA), and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense (coding) or antisense (non-coding) strand. By way of nonlimiting example, fragments include nucleic acid sequences that can be about 10 to 60 contiguous nucleotides in

length, preferably, at least 15-60 contiguous nucleotides in length, and also preferably include fragments that are at least 70-100 contiguous nucleotides, or which are at least 1000 contiguous nucleotides or greater in length. Nucleic acids for use as probes or primers may differ in length as described herein.

In specific embodiments, HDAC polynucleotides of the present invention can comprise at least 15, 20, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1195, 1200, 1500, 2000, 2160, 2250, 2500, 2755, or 2900 contiguous nucleotides of SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, SEQ ID NO:96, or a sequence complementary thereto. Additionally, a polynucleotide of the invention can comprise a specific region of a HDAC nucleotide sequence, e.g., a region encoding the C-terminal sequence of the HDAC polypeptide. Such polynucleotides can comprise, for example, nucleotides 3024-4467 of HDAC9c (SEQ ID NO:88), nucleotides 2156-3650 of HDACX_v1 (SEQ ID NO:94), nucleotides 1174-3391 of HDACX_v2 (SEQ ID NO:96), or portions or fragments thereof.

As specific examples, polynucleotides of the invention may comprise at least 183 contiguous nucleotides of SEQ ID NO:88; or at least 17 contiguous nucleotides of SEQ ID NO:96. As additional examples, the polynucleotides of the invention may comprise nucleotides 1 to 3207 of SEQ ID NO:88; nucleotides 1 to 2340 of SEQ ID NO:94; or nucleotides 307 to 1791 of SEQ ID NO:96. Further, the polynucleotides of the invention may comprise nucleotides 4 to 3207 of SEQ ID NO:88, wherein said nucleotides encode amino acids 2 to 1069 of SEQ ID NO:87 lacking the start methionine; or nucleotides 310 to 1791 of SEQ ID NO:96, wherein said nucleotides encode amino acids 2 to 495 of SEQ ID NO:95 lacking the start methionine. In addition, polynucleotides of the invention may comprise nucleotides 3024-3207 of SEQ ID NO:88; or nucleotides 1174-1791 of SEQ ID NO:96.

"Amino acid sequence" as used herein refers to an oligopeptide, peptide, polypeptide, or protein sequence, and fragments or portions thereof, and to naturally occurring or synthetic molecules. Amino acid sequence

fragments are typically from about 4 or 5 to about 35, preferably from about 5 to about 15 or 25 amino acids in length and, optimally, retain the biological activity or function of an HDAC polypeptide. However, it will be understood that larger amino acid fragments can be used, depending on the purpose therefor, e.g., fragments of from about 15 to about 50 or 60 amino acids, or greater.

Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms, such as "polypeptide" or "protein" are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule. In addition, the terms HDAC polypeptide and HDAC protein are frequently used interchangeably herein to refer to the encoded product of an HDAC nucleic acid sequence of the present invention.

A variant of an HDAC polypeptide can refer to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant may have "nonconservative" changes, e.g., replacement of a glycine with a tryptophan. Minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing functional biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

An allele or allelic sequence is an alternative form of an HDAC nucleic acid sequence. Alleles may result from at least one mutation in the nucleic acid sequence and may yield altered mRNAs or polypeptides whose structure or function may or may not be altered. Any given gene, whether natural or recombinant, may have none, one, or many allelic forms. Common mutational changes that give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of

changes may occur alone, or in combination with the others, one or more times in a given sequence.

Altered nucleic acid sequences encoding an HDAC polypeptide include nucleic acid sequences containing deletions, insertions and/or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent HDAC polypeptide. Altered nucleic acid sequences may further include polymorphisms of the polynucleotide encoding an HDAC polypeptide; such polymorphisms may or may not be readily detectable using a particular oligonucleotide probe. The encoded protein may also contain deletions, insertions, or substitutions of amino acid residues, which produce a silent change and result in a functionally equivalent HDAC protein of the present invention. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological activity or function of the HDAC protein is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid; positively charged amino acids may include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide ("oligo") linked to a peptide backbone of amino acid residues, which terminates in lysine. PNA typically comprise oligos of at least 5 nucleotides linked to amino acid residues. These small molecules stop transcript elongation by binding to their complementary strand of nucleic acid (P.E. Nielsen et al., 1993, *Anticancer Drug Des.*, 8:53-63). PNA may be pegylated to extend their lifespan in the cell where they preferentially bind to complementary single stranded DNA and RNA.

Oligonucleotides or oligomers refer to a nucleic acid sequence, preferably comprising contiguous nucleotides, typically of at least about 6 nucleotides to about 60 nucleotides, preferably at least about 8 to 10 nucleotides in length, more preferably at least about 12 nucleotides in length,

e.g., about 15 to 35 nucleotides, or about 15 to 25 nucleotides, or about 20 to 35 nucleotides, which can be typically used, for example, as probes or primers, in PCR amplification assays, hybridization assays, or in microarrays. It will be understood that the term oligonucleotide is substantially equivalent to the terms primer, probe, or amplimer, as commonly defined in the art. It will also be appreciated by those skilled in the pertinent art that a longer oligonucleotide probe, or mixtures of probes, e.g., degenerate probes, can be used to detect longer, or more complex, nucleic acid sequences, for example, genomic DNA. In such cases, the probe may comprise at least 20-200 nucleotides, preferably, at least 30-100 nucleotides, more preferably, 50-100 nucleotides.

Amplification refers to the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction (PCR) technologies, which are well known and practiced in the art (See, D.W. Dieffenbach and G.S. Dveksler, 1995, *PCR Primer, a Laboratory Manual,* Cold Spring Harbor Press, Plainview, NY).

Microarray is an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon, or other type of membrane; filter; chip; glass slide; or any other type of suitable solid support.

The term antisense refers to nucleotide sequences, and compositions containing nucleic acid sequences, which are complementary to a specific DNA or RNA sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense (i.e., complementary) nucleic acid molecules include PNA and may be produced by any method, including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes that block either transcription or translation. The designation "negative" is sometimes used in reference to the antisense strand, and "positive" is sometimes used in reference to the sense strand.

The term consensus refers to the sequence that reflects the most common choice of base or amino acid at each position among a series of

related DNA, RNA, or protein sequences. Areas of particularly good agreement often represent conserved functional domains.

A deletion refers to a change in either nucleotide or amino acid sequence and results in the absence of one or more nucleotides or amino acid residues. By contrast, an insertion (also termed "addition") refers to a change in a nucleotide or amino acid sequence that results in the addition of one or more nucleotides or amino acid residues, as compared with the naturally occurring molecule. A substitution refers to the replacement of one or more nucleotides or amino acids by different nucleotides or amino acids.

A derivative nucleic acid molecule refers to the chemical modification of a nucleic acid encoding, or complementary to, an encoded HDAC polypeptide. Such modifications include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative encodes a polypeptide that retains the essential biological and/or functional characteristics of the natural molecule. A derivative polypeptide is one that is modified by glycosylation, pegylation, or any similar process that retains the biological and/or functional or immunological activity of the polypeptide from which it is derived.

The term "biologically active", i.e., functional, refers to a protein or polypeptide or peptide fragment thereof having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic HDAC, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells, for example, to generate antibodies, and to bind with specific antibodies.

An HDAC-related protein refers to the HDAC and HADL proteins or polypeptides described herein, as well as other human homologs of these HDAC or HDAL sequences, in addition to orthologs and paralogs (homologs) of the HDAC or HADL sequences in other species, ranging from yeast to other mammals, e.g., homologous histone deacetylase. The term ortholog refers to genes or proteins that are homologs via speciation, e.g., closely related and assumed to have common descent based on structural and

functional considerations. Orthologous proteins function as recognizably the same activity in different species. The term paralog refers to genes or proteins that are homologs via gene duplication, e.g., duplicated variants of a gene within a genome. (See, W.M. Fritch, 1970, *Syst. Zool.*, 19:99-113.

It will be appreciated that, under certain circumstances, it may be advantageous to provide homologs of one of the novel HDAC polypeptides which function in a limited capacity as one of either an HDAC agonist (i.e., mimetic), or an HDAC antagonist, in order to promote or inhibit only a subset of the biological activities of the naturally-occurring form of the protein. Thus, specific biological effects can be elicited by treatment with a homolog of limited function, and with fewer side effects, relative to treatment with agonists or antagonists which are directed to all of the biological activities of naturally-occurring forms of HDAC proteins.

Homologs (i.e., isoforms or variants) of the novel HDAC polypeptides can be generated by mutagenesis, such as by discrete point mutation(s), or by truncation. For example, mutation can yield homologs that retain substantially the same, or merely a subset of, the biological activity of the HDAC polypeptide from which it was derived. Alternatively, antagonistic forms of the protein can be generated which are able to inhibit the function of the naturally-occurring form of the protein, such as by competitively binding to an HDAC substrate, or HDAC-associated protein. Non-limiting examples of such situations include competing with wild-type HDAC in the binding of p53 or a histone. Also, agonistic forms of the protein can be generated which are constitutively active, or have an altered K_{cat} or K_m for deacylation reactions. Thus, the HDAC protein and homologs thereof may be either positive or negative regulators of transcription and/or replication.

The term hybridization refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary G and C bases and between complementary A and T bases. The hydrogen bonds may be further stabilized by base stacking

interactions. The two complementary nucleic acid sequences hydrogen bond in an anti-parallel configuration. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis), or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., membranes, filters, chips, pins, or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been affixed).

The terms stringency or stringent conditions refer to the conditions for hybridization as defined by nucleic acid composition, salt and temperature. These conditions are well known in the art and may be altered to identify and/or detect identical or related polynucleotide sequences in a sample. A variety of equivalent conditions comprising either low, moderate, or high stringency depend on factors such as the length and nature of the sequence (DNA, RNA, base composition), reaction milieu (in solution or immobilized on a solid substrate), nature of the target nucleic acid (DNA, RNA, base composition), concentration of salts and the presence or absence of other reaction components (e.g., formamide, dextran sulfate and/or polyethylene glycol) and reaction temperature (within a range of from about 5°C below the melting temperature of the probe to about 20°C to 25°C below the melting temperature). One or more factors may be varied to generate conditions, either low or high stringency, that are different from but equivalent to the aforementioned conditions.

As will be understood by those of skill in the art, the stringency of hybridization may be altered in order to identify or detect identical or related polynucleotide sequences. As will be further appreciated by the skilled practitioner, Tm can be approximated by the formulas as known in the art, depending on a number of parameters, such as the length of the hybrid or probe in number of nucleotides, or hybridization buffer ingredients and conditions (See, for example, T. Maniatis et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982 and J. Sambrook et al., *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989; *Current Protocols in Molecular Biology*, Eds. F.M. Ausubel et al., Vol. 1, "Preparation and Analysis

of DNA", John Wiley and Sons, Inc., 1994-1995, Suppls. 26, 29, 35 and 42; pp. 2.10.7- 2.10.16; G.M. Wahl and S. L. Berger (1987; *Methods Enzymol.* 152:399-407); and A.R. Kimmel, 1987; *Methods of Enzymol.*, 152:507-511). As a general guide, Tm decreases approximately 1ºC –1.5ºC with every 1% decrease in sequence homology. Also, in general, the stability of a hybrid is a function of sodium ion concentration and temperature. Typically, the hybridization reaction is initially performed under conditions of low stringency, followed by washes of varying, but higher stringency. Reference to hybridization stringency, e.g., high, moderate, or low stringency, typically relates to such washing conditions.

Thus, by way of nonlimiting example, high stringency refers to conditions that permit hybridization of those nucleic acid sequences that form stable hybrids in 0.018M NaCl at about 65ºC (i.e., if a hybrid is not stable in 0.018M NaCl at about 65ºC, it will not be stable under high stringency conditions). High stringency conditions can be provided, for instance, by hybridization in 50% formamide, 5 X Denhart's solution, 5 X SSPE (saline sodium phosphate EDTA) (1 X SSPE buffer comprises 0.15 M NaCl, 10 mM Na_2HPO_4, 1 mM EDTA), (or 1 X SSC buffer containing 150 mM NaCl, 15 mM Na_3 citrate • 2 H_2O, pH 7.0), 0.2% SDS at about 42ºC, followed by washing in 1 X SSPE (or saline sodium citrate, SSC) and 0.1% SDS at a temperature of at least about 42°C, preferably about 55°C, more preferably about 65°C.

Moderate stringency refers, by way of nonlimiting example, to conditions that permit hybridization in 50% formamide, 5 X Denhart's solution, 5 X SSPE (or SSC), 0.2% SDS at 42ºC (to about 50ºC), followed by washing in 0.2 X SSPE (or SSC) and 0.2% SDS at a temperature of at least about 42°C, preferably about 55°C, more preferably about 65°C.

Low stringency refers, by way of nonlimiting example, to conditions that permit hybridization in 10% formamide, 5 X Denhart's solution, 6 X SSPE (or SSC), 0.2% SDS at 42ºC, followed by washing in 1 X SSPE (or SSC) and 0.2% SDS at a temperature of about 45°C, preferably about 50°C.

For additional stringency conditions, see T. Maniatis et al., *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring

Harbor, NY (1982). It is to be understood that the low, moderate and high stringency hybridization / washing conditions may be varied using a variety of ingredients, buffers and temperatures well known to and practiced by the skilled practitioner.

The terms complementary or complementarity refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A". Complementarity between two single-stranded molecules may be "partial", in which only some of the nucleic acids bind, or it may be complete when total complementarity exists between single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands, as well as in the design and use of PNA molecules.

The term homology refers to a degree of complementarity. There may be partial sequence homology or complete homology, wherein complete homology is equivalent to identity, e.g., 100% identity. A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (e.g., Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous sequence or probe to the target sequence under conditions of low stringency. Nonetheless, conditions of low stringency do not permit non-specific binding; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% identity). In the absence of non-

specific binding, the probe will not hybridize to the second non-complementary target sequence.

Those having skill in the art will know how to determine percent identity between/among sequences using, for example, algorithms such as those based on the CLUSTALW computer program (J.D. Thompson et al., 1994, *Nucleic Acids Research*, 2(22):4673-4680), or FASTDB, (Brutlag et al., 1990, *Comp. App. Biosci.*, 6:237-245), as known in the art. Although the FASTDB algorithm typically does not consider internal non-matching deletions or additions in sequences, i.e., gaps, in its calculation, this can be corrected manually to avoid an overestimation of the % identity. CLUSTALW, however, does take sequence gaps into account in its identity calculations.

Also available to those having skill in this art are the BLAST and BLAST 2.0 algorithms (Altschul et al., 1977, *Nucl. Acids Res.*, 25:3389-3402 and Altschul et al., 1990, *J. Mol. Biol.*, 215:403-410). The BLASTN program for nucleic acid sequences uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, and an expectation (E) of 10. The BLOSUM62 scoring matrix (Henikoff and Henikoff, 1989, *Proc. Natl. Acad. Sci., USA*, 89:10915) uses alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

An HDAC polynucleotide of the present invention may show at least 27.7%, 35%, 40%, 44.1%, 48.2%, 50%, 55.4%, 58.6%, 59.8%, 60%, 60.2%, 67.8%, 70%, 80%, 81.5%, 85%, 90%, 91%, 92%, 93%, 94%, 94.2%, 94.4%, 95%, 96%, 97%, 97.2%, 97.5%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, or 99.9% identity to a sequence provided in SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, SEQ ID NO:96, or a sequence complementary thereto. An HDAC polypeptide of the present invention may show at least 25%, 35%, 40%, 45%, 48.1%, 55.2%, 55.3%, 60%, 65%, 70%, 72%, 75%, 79%, 80%, 80.6%, 85%, 90%, 91%, 92%, 93%, 94%, 94.2%, 95%, 96%, 97%, 97.2%, 97.5%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, or 99.9%

identity to a sequence provided in any one of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, or SEQ ID NO:95.

In a preferred aspect of the invention, a HDAC polynucleotide shows at least 60.2%, 81.5%, or 94.4% identity to the HDAC9c nucleotide sequence (SEQ ID NO:88 or a sequence complementary thereto); or at least 27.7%, 48.2%, or 55.4% identity to the HDACX_v2 nucleotide sequence (SEQ ID NO:96 or a sequence complementary thereto). A HDAC polypeptide of the invention preferably shows at least 55.2%, 80.6%, or 94.2% identity to the HDAC9c amino acid sequence (SEQ ID NO:87); at least 55.3% identity to the HDACX_v2 amino acid sequence (SEQ ID NO:95); at least 72% identity to the amino acid sequence of BMY_HDAL1 (SEQ ID NO:2); at least 79% identity to the amino acid sequence of BMY_HDAL2 (SEQ ID NO:4); or at least 70% identity to the amino acid sequence of BMY_HDAL3 (SEQ ID NO:5).

A composition comprising a given polynucleotide sequence refers broadly to any composition containing the given polynucleotide sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising the polynucleotide sequences (e.g., SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, or SEQ ID NO:96) encoding the novel HDAC polypeptides of this invention, or fragments thereof, or complementary sequences thereto, may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be in association with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be employed in an aqueous solution containing salts (e.g., NaCl), detergents or surfactants (e.g., SDS) and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, and the like).

The term "substantially purified" refers to nucleic acid sequences or amino acid sequences that are removed from their natural environment, i.e., isolated or separated by a variety of means, and are at least 60% free, preferably 75% to 85% free, and most preferably 90% or greater free from other components with which they are naturally associated.

The term sample, or biological sample, is meant to be interpreted in its broadest sense. A biological sample suspected of containing nucleic acid encoding an HDAC protein, or fragments thereof, or an HDAC protein itself, may comprise a body fluid, an extract from cells or tissue, chromosomes isolated from a cell (e.g., a spread of metaphase chromosomes), organelle, or membrane isolated from a cell, a cell, nucleic acid such as genomic DNA (in solution or bound to a solid support such as for Southern analysis), RNA (in solution or bound to a solid support such as for Northern analysis), cDNA (in solution or bound to a solid support), a tissue, a tissue print and the like.

Transformation refers to a process by which exogenous DNA enters and changes a recipient cell. It may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and partial bombardment. Such "transformed" cells include stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome. Transformed cells also include those cells that transiently express the inserted DNA or RNA for limited periods of time.

The term "mimetic" refers to a molecule, the structure of which is developed from knowledge of the structure of an HDAC protein, or portions thereof, and as such, is able to effect some or all of the actions of HDAC proteins.

The term "portion" with regard to a protein (as in "a portion of a given protein") refers to fragments or segments, for example, peptides, of that protein. The fragments may range in size from four or five amino acid residues to the entire amino acid sequence minus one amino acid. Thus, a protein "comprising at least a portion of the amino acid sequence of the HDAC molecules presented herein can encompass a full-length human HDAC polypeptide, and fragments thereof.

In specific embodiments, HDAC polypeptides of the invention can comprise at least 5, 10, 20, 30, 50, 70, 100, 200, 300, 400, 500, 600, 700, 720, 750, 800, 920, or 950 contiguous amino acid residues of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, or SEQ ID NO:95. Additionally, a polypeptide of the invention can comprise a specific region, e.g., the C-terminal region, of a HDAC amino acid sequence. Such polypeptides can comprise, for example, amino acids 1009-1069 of HDAC9c (SEQ ID NO:87), amino acids 720-780 of HDACX_v1 (SEQ ID NO:93), or portions or fragments thereof.

The term antibody refers to intact molecules as well as fragments thereof, such as Fab, $F(ab')_2$, Fv, which are capable of binding an epitopic or antigenic determinant. Antibodies that bind to the HDAC polypeptides can be prepared using intact polypeptides or fragments containing small peptides of interest or prepared recombinantly for use as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal can be derived from the transition of RNA or synthesized chemically, and can be conjugated to a carrier protein, if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin (BSA), keyhole limpet hemocyanin (KLH), and thyroglobulin. The coupled peptide is then used to immunize the animal (e.g, a mouse, a rat, or a rabbit).

The term "humanized" antibody refers to antibody molecules in which amino acids have been replaced in the non-antigen binding regions, e.g., the complementarity determining regions (CDRs), in order to more closely resemble a human antibody, while still retaining the original binding capability, e.g., as described in U.S. Patent No. 5,585,089 to C.L. Queen et al., which is a nonlimiting example. Fully humanized antibodies, such as those produced transgenically or recombinantly, are also encompassed herein.

The term "antigenic determinant" refers to that portion of a molecule that makes contact with a particular antibody (i.e., an epitope). When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein;

these regions or structures are referred to an antigenic determinants. An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The terms "specific binding" or "specifically binding" refer to the interaction between a protein or peptide and a binding molecule, such as an agonist, an antagonist, or an antibody. The interaction is dependent upon the presence of a particular structure (e.g., an antigenic determinant or epitope, or a structural determinant) of the protein that is recognized by the binding molecule. For example, if an antibody is specific for epitope "A", the presence of a protein containing epitope A (or free, unlabeled A) in a reaction containing labeled "A" and the antibody will reduce the amount of labeled A bound to the antibody.

The term "correlates with expression of a polynucleotide" indicates that the detection of the presence of ribonucleic acid that is similar to one or more of the HDAC sequences provided herein by Northern analysis is indicative of the presence of mRNA encoding an HDAC polypeptide in a sample and thereby correlates with expression of the transcript from the polynucleotide encoding the protein.

An alteration in the polynucleotide of an HDAC nucleic acid sequence comprises any alteration in the sequence of the polynucleotides encoding an HDAC polypeptide, including deletions, insertions, and point mutations that may be detected using hybridization assays. Included within this definition is the detection of alterations to the genomic DNA sequence which encodes an HDAC polypeptide (e.g., by alterations in the pattern of restriction fragment length polymorphisms capable of hybridizing to the HDAC nucleic acid sequences presented herein, (i.e., SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, and/or SEQ ID NO:96), the inability of a selected fragment of a given HDAC sequence to hybridize to a sample of genomic DNA (e.g., using allele-specific oligonucleotide probes), and improper or unexpected hybridization, such as hybridization to a locus other than the normal chromosomal locus for the polynucleotide sequence encoding

an HDAC polypeptide (e.g., using fluorescent *in situ* hybridization (FISH) to metaphase chromosome spreads).

Description of Embodiments of the Present Invention

In one of its embodiments, the present invention is directed to a novel HDAC termed, BMY_HDAL1, which is encoded by the human BAC clones AC016186, AC00755 and AC002088. The BMY_HDAL1 nucleic acid (cDNA) sequence is provided as SEQ ID NO:1; the BMY_HDAL1 amino acid sequence encoded by the BMY_HDAL1 nucleic acid sequence is presented as SEQ ID NO:2. (FIG. 1).

BMY_HDAL1 was identified by HMM analysis using PFAM model PF00850. (Example 1). The PFAM-HMM database is a collection of protein families and domains and contains multiple protein alignments (A. Bateman et al., 1999, *Nucleic Acids Research*, 27:260-262). BMY_HDAL1 is most closely related to the known human histone deacetylase HDAC5; the two proteins are 71% identical and 77% similar over 105 amino acids, as determined by the GCG Gap program with a gap weight of 8 and a length weight of 2. The gene structure and predicted cDNA and protein sequence of BMY_HDAL1 were determined by comparison to the known human histone deacetylase HDAC5 using the GenewiseDB program to analyze human BAC AC002088 (E. Birney and R. Durbin, 2000, *Genome Res.*, 10(4):547-548).

Sequence motifs of BMY_HDAL1 were examined using the GCG Motifs program to ascertain if there were motifs common to other known proteins in the PROSITE collection (K. Hofmann et al., 1999, *Nucleic Acids Res.*, 27(1):215-219) with no allowed mismatches. Motifs programs typically search for protein motifs by searching protein sequences for regular-expression patterns described in the PROSITE Dictionary. FIG. 4 shows PROSITE motifs identified in the partial predicted amino acid sequence of BMY_HDAL1.

In another embodiment, the present invention is directed to the novel HDAC termed BMY_HDAL2, a novel human histone deacetylase-like protein encoded by genomic BACs AC002410. The BMY_HDAL2 nucleic acid sequence (SEQ ID NO:12) and its encoded polypeptide (SEQ ID NO:4) are

presented in FIG. 5. BMY_HDAL2 was identified by hidden Markov model searches using the PFAM HMM PF00850 to search predicted proteins from human genomic DNA. BMY_HDAL2 is most closely related to the known human histone deacetylase HDAC5; the two proteins are 78% identical and 86% similar over 163 amino acids as determined by the GCG Gap program with a gap weight of 8 and a length weight of 2. The gene structure and predicted cDNA and protein sequences of BMY_HDAL2 were determined by comparison to BMY_HDA5 using the GenewiseDB program (E. Birney and R. Durbin, 2000, *Genome Res.*, 10(4):547-548).

Sequence motifs of BMY_HDAL2 were examined using the GCG Motifs program to ascertain if there were motifs in the PROSITE collection (K. Hofmann et al., 1999, *Nucleic Acids Res.*, 27(1):215-219) with no allowed mismatches. FIG. 7 shows PROSITE motifs identified in the partial predicted amino acid sequence of BMY_HDAL2.

In addition, the genomic location surrounding BMY_HDAL2 was investigated. Based on the genomic location of BAC AC002410 as reported by the NCBI MapViewer, BMY_HDAL2 has been localized to chromosome 7 region q36.

In another embodiment, the present invention further provides a third HDAC termed BMY_HDAL3. The BMY_HDAL3 nucleic acid sequence (SEQ ID NO:19) and its encoded polypeptide (SEQ ID NO:5) are presented in FIG. 10. BMY_HDAL3 is encoded by the human genomic BAC clones AC004994 and AC004744. BMY_HDAL3 was identified by HMM analysis using PFAM model PF00850 to search predicted proteins generated from human genomic DNA sequences using Genscan. BMY_HDAL3 is most closely related to the known human histone deacetylase HDAC5; the two proteins are 69% identical over 1122 amino acids as determined by the GCG Gap program with a gap weight of 8 and a length weight of 2.

The partial transcripts identified from BAC clones AC004994 (SEQ ID NO:15) and AC004744 (SEQ ID NO:16) were assembled into a single contig (designated BMY_HDAL3) using the VectorNTI ContigExpress program (Informax). (FIG. 9). The gene structure and predicted cDNA and protein

sequence of BMY_HDAL3 were determined by comparison to the known human histone deacetylase HDAC5 using the GenewiseDB program (K. Hofmann et al., 1999, *Nucleic Acids Res.*, 27(1):215-219) and are presented in FIG. 9. The most N-terminal region of the BMY_HDAL3 sequence described herein is encoded by human genomic BAC AC004994. (FIG. 8A).

BMY_HDAL3 has been localized to chromosome 7, region q36 based on the locations reported for AC004994 and by the NCBI MapViewer.

Sequence motifs of BMY_HDAL3 were examined using the GCG Motifs program to ascertain if there were motifs in the PROSITE collection (K. Hofmann et al., 1999, *Nucleic Acids Res.*, 27(1):215-219) with no allowed mismatches. FIG. 11 shows PROSITE motifs identified in the partial predicted amino acid sequence of BMY_HDAL3. FIG. 12 shows a multiple sequence alignment of the novel human HDAC, BMY_HDAL3, amino acid sequence (SEQ ID NO:5) with the amino acid sequence of AAC78618 (SEQ ID NO:21) and with the amino acid sequence of AAD15364 (SEQ ID NO:22). AAC78618 is a histone deacetylase-like protein predicted by genefinding and conceptual translation of AC004994 and which was entered in Genbank. AAD15364 is a similar predicted protein derived from AC004744 and entered in Genbank. AAC78618, AAD15364 and BMY_HDAL3 were aligned using the ClustalW algorithm as implemented in the VectorNTI sequence analysis package (1998, 5.5 Ed., Informax, Inc.) with a gap opening penalty of 10, a gap extension penalty of 0.1 and no end gap penalties.

Novel HDAC9 variants, termed HDACX_v1 and HDACX_v2, have also been identified. In addition, HDAC9c, an HDAC9-related family member, has been newly identified and cloned.

<u>HDAC Polynucleotides and Polypeptides</u>

The present invention encompasses novel HDAC nucleic acid sequences (e.g., SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, SEQ ID NO:96, and sequences complementary thereto) encoding newly discovered histone deacetylase like polypeptides (e.g., SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, and SEQ ID NO:95). These HDAC polynucleotides, polypeptides, or

compositions thereof, can be used in methods for screening for antagonists or inhibitors of the activity or function of HDACs.

In another of its embodiments, the present invention encompasses new HDAC polypeptides comprising the amino acid sequences of, e.g., SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, and SEQ ID NO:95, and as shown in FIG. 1, FIG. 5, FIG. 10, FIGS. 15A-15C, FIGS. 20A-20C, and FIGS. 21A-21B.

The HDAC polypeptides as described herein show close similarity to HDAC proteins, including HDAC5 and HDAC9. FIGS. 2A and 2B portray the structural similarities among the novel HDAC polypeptides and several other proteins, namely Aquifex HDAL, Human HDAC4, Human HDAC5, Human HDAC7, and *Saccharomyces cerevisiae* HDA1. FIGS. 15D-15F show the amino acid sequence similarity and identity shared by HDAC9c and previously identified HDAC9 amino acid sequences. FIGS. 23A-23K show the nucleotide sequence identity shared by HDACX_v1, HDACX_v2, and previously identified HDAC9 nucleotide sequences.

Variants of the disclosed HDAC polynucleotides and polypeptides are also encompassed by the present invention. In some cases, a HDAC polynucleotide variant (i.e., variant of SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, or SEQ ID NO:96) will encode an amino acid sequence identical to a HDAC sequence (e.g., SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, and SEQ ID NO:95). This is due to the redundancy (degeneracy) of the genetic code, which allows for silent mutations. In other cases, a HDAC polynucleotide variant will encode a HDAC polypeptide variant (i.e., a variant of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, or SEQ ID NO:95). Preferably, an HDAC polypeptide variant has at least 75 to 80%, more preferably at least 85 to 90%, and even more preferably at least 90% or greater amino acid sequence identity to one or more of the HDAC amino acid sequences (e.g., SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, and SEQ ID NO:95) as disclosed herein, and which retains at least one biological or other functional characteristic or activity of the HDAC

polypeptide. Most preferred is a variant having at least 95% amino acid sequence identity to the amino acid sequences set forth in SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, and SEQ ID NO:95.

An amino acid sequence variant of the HDAC proteins can be categorized into one or more of three classes: substitutional, insertional, or deletional variants. Such variants are typically prepared by site-specific mutagenesis of nucleotides in the DNA encoding the HDAC protein, using cassette or PCR mutagenesis, or other techniques that are well known and practiced in the art, to produce DNA encoding the variant. Thereafter, the DNA is expressed in recombinant cell culture as described herein. Variant HDAC protein fragments having up to about 100-150 residues may be prepared by in vitro synthesis using conventional techniques.

Amino acid sequence variants are characterized by the predetermined nature of the variation, a feature that sets them apart from naturally occurring allelic or interspecies variations of an HDAC amino acid sequence. The variants typically exhibit the same qualitative biological activity as that of the naturally occurring analogue, although variants can also be selected having modified characteristics. While the site or region for introducing an amino acid sequence variation is predetermined, the mutation per se need not be predetermined. For example, in order to optimize the performance of a mutation at a given site, random mutagenesis may be performed at the target codon or region, and the expressed HDAC variants can be screened for the optimal combination of desired activity. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence are well known, for example, M13 primer mutagenesis and PCR mutagenesis. Screening of the mutants is accomplished using assays of HDAC protein activity, for example, for binding domain mutations, competitive binding studies may be carried out.

Amino acid substitutions are typically of single residues; insertions usually are on the order of from one to twenty amino acids, although considerably larger insertions may be tolerated. Deletions range from about

one to about 20 residues, although in some cases, deletions may be much larger.

Substitutions, deletions, insertions, or any combination thereof, may be used to arrive at a final HDAC derivative. Generally, these changes affect only a few amino acids to minimize the alteration of the molecule. However, larger changes may be tolerated in certain circumstances. When small alterations in the characteristics of the HDAC protein are desired or warranted, substitutions are generally made in accordance with the following table:

Original Residue	Conservative Substitution(s)	Original Residue	Conservative Substitution(s)
Ala	Ser	Leu	Ile, Val
Arg	Lys	Lys	Arg, Gln, Glu
Asn	Gln, His	Met	Leu, Ile
Asp	Glu	Phe	Met, Leu, Tyr
Cys	Ser	Ser	Thr
Gln	Asn	Thr	Ser
Glu	Asp	Trp	Tyr
Gly	Pro	Tyr	Trp, Phe
His	Asn, Gln	Val	Ile, Leu
Ile	Leu, Val		

Substantial changes in function or immunological identity are made by selecting substitutions that are less conservative than those shown in the above Table. For example, substitutions may be made which more significantly affect the structure of the polypeptide backbone in the area of the alteration, for example, the alpha-helical, or beta-sheet structure; the charge or hydrophobicity of the molecule at the target site; or the bulk of the side chain. The substitutions which generally are expected to produce the greatest changes in the polypeptide's properties are those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl, or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue

having a bulky side chain, e.g., phenylalanine, is substituted for (or by) a residue that does not have a side chain, e.g., glycine.

While HDAC variants will ordinarily exhibit the same qualitative biological activity or function, and elicit the same immune response, as the naturally occurring analogue, the variants are also selected to modify the characteristics of HDAC proteins as needed. Alternatively, the variant may be designed such the that biological activity of the HDAC protein is altered, e.g., improved.

In another embodiment, the present invention encompasses polynucleotides that encode the novel HDAC polypeptides disclosed herein. Accordingly, any nucleic acid sequence that encodes the amino acid sequence of an HDAC polypeptide of the invention can be used to produce recombinant molecules that express that HDAC protein. In a particular embodiment, the present invention encompasses the novel human HDAC polynucleotides comprising the nucleic acid sequences of SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, and SEQ ID NO:96 as shown in FIG. 1, FIG. 5, FIG. 10, FIGS. 15A-15C, FIGS. 20A-20C, and FIGS. 21A-21B. More particularly, the present invention embraces cloned full-length open reading frame human BMY_HDAL1, BMY_HDAL2 and BMY_HDAL3 deposited at the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209 on ______________________________ under ATCC Accession No. ____ ____________________________ according to the terms of the Budapest Treaty.

As will be appreciated by the skilled practitioner in the art, the degeneracy of the genetic code results in the production of more than one appropriate nucleotide sequence encoding the HDAC polypeptides of the present invention. Some of the sequences bear minimal homology to the nucleotide sequences of any known and naturally occurring gene. Accordingly, the present invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are

made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of a naturally occurring HDAC protein, and all such variations are to be considered as being embraced herein.

Although nucleotide sequences which encode the HDAC polypeptides and variants thereof are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring HDAC polypeptides under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding the HDAC polypeptides, or derivatives thereof, which possess a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide/polypeptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host, for example, in plant cells or yeast cells or amphibian cells. Other reasons for substantially altering the nucleotide sequence encoding the HDAC polypeptides, and derivatives, without altering the encoded amino acid sequences, include the production of mRNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The present invention also encompasses production of DNA sequences, or portions thereof, which encode the HDAC polypeptides, and derivatives of these polypeptides, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known and practiced by those in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding an HDAC polypeptide, or any fragment thereof.

Also encompassed by the present invention are polynucleotide sequences that are capable of hybridizing to the HDAC nucleotide sequences presented herein, such as those shown in SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, and SEQ ID NO:96, or sequences complementary thereto, under various conditions of stringency. Hybridization conditions are typically based on the melting temperature (Tm) of the nucleic

acid binding complex or probe (See, G.M. Wahl and S.L. Berger, 1987; *Methods Enzymol.*, 152:399-407 and A.R. Kimmel, 1987; *Methods of Enzymol.*, 152:507-511), and may be used at a defined stringency. For example, included in the present invention are sequences capable of hybridizing under moderately stringent conditions to the HDAC nucleic acid sequences of SEQ ID NO:1, SEQ ID NO:12, or SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, and SEQ ID NO:96, and other sequences which are degenerate to those which encode the HDAC polypeptides (e.g., as a nonlimiting example: prewashing solution of 2 X SSC, 0.5% SDS, 1.0mM EDTA, pH 8.0, and hybridization conditions of 50°C, 5 X SSC, overnight).

In another embodiment of the present invention, polynucleotide sequences or fragments (peptides) thereof which encode the HDAC polypeptide may be used in recombinant DNA molecules to direct the expression of the HDAC polypeptide products, or fragments or functional equivalents thereof, in appropriate host cells. Because of the inherent degeneracy of the genetic code, other DNA sequences, which encode substantially the same or a functionally equivalent amino acid sequences, may be produced, and these sequences may be used to express recombinant HDAC polypeptides.

As will be appreciated by those having skill in the art, it may be advantageous to produce HDAC polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter HDAC polypeptide-encoding sequences for a variety of reasons, including, but not limited to, alterations which modify the cloning, processing, and/or expression of the gene products. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer

the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and the like.

In another embodiment of the present invention, natural, modified, or recombinant nucleic acid sequences, or a fragment thereof, encoding the HDAC polypeptides may be ligated to a heterologous sequence to encode a fusion protein. For example, for screening peptide libraries for inhibitors or modulators of HDAC activity or binding, it may be useful to encode a chimeric HDAC protein or peptide that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between an HDAC protein-encoding sequence and the heterologous protein sequence, so that the HDAC protein may be cleaved and purified away from the heterologous moiety.

In another embodiment, ligand-binding assays are useful to identify inhibitor or antagonist compounds that interfere with the function of the HDAC protein, or activator compounds that stimulate the function of the HDAC protein. Preferred are inhibitor or antagonist compounds. Such assays are useful even if the function of a protein is not known. These assays are designed to detect binding of test compounds (i.e., test agents) to particular target molecules, e.g., proteins or peptides. The detection may involve direct measurement of binding. Alternatively, indirect indications of binding may involve stabilization of protein structure, or disruption or enhancement of a biological function. Non-limiting examples of useful ligand-binding assays are detailed below.

One useful method for the detection and isolation of binding proteins is the Biomolecular Interaction Assay (BIAcore) system developed by Pharmacia Biosensor and described in the manufacturer's protocol (LKB Pharmacia, Sweden). The BIAcore system uses an affinity purified anti-GST antibody to immobilize GST-fusion proteins onto a sensor chip. The sensor utilizes surface plasmon resonance, which is an optical phenomenon that detects changes in refractive indices. Accordingly, a protein of interest, e.g., an HDAC polypeptide, or fragment thereof, of the present invention, is coated

onto a chip and test compounds (i.e., test agents) are passed over the chip. Binding is detected by a change in the refractive index (surface plasmon resonance).

A different type of ligand-binding assay involves scintillation proximity assays (SPA), as described in U.S. Patent No. 4,568,649. In a modification of this assay currently undergoing development, chaperonins are used to distinguish folded and unfolded proteins. A tagged protein is attached to SPA beads, and test compounds are added. The bead is then subjected to mild denaturing conditions, such as, for example, heat, exposure to SDS, and the like, and a purified labeled chaperonin is added. If a test compound (i.e., test agent) has bound to a target protein, the labeled chaperonin will not bind; conversely, if no test compound has bound, the protein will undergo some degree of denaturation and the chaperonin will bind. In another type of ligand binding assay, proteins containing mitochondrial targeting signals are imported into isolated mitochondria *in vitro* (Hurt et al., 1985, *EMBO J.*, 4:2061-2068; Eilers and Schatz, 1986, *Nature*, 322:228-231).

In a mitochondrial import assay, expression vectors are constructed in which nucleic acids encoding particular target proteins are inserted downstream of sequences encoding mitochondrial import signals. The chimeric proteins are synthesized and tested for their ability to be imported into isolated mitochondria in the absence and presence of test compounds. A test compound that binds to the target protein should inhibit its uptake into isolated mitochondria *in vitro*.

Another type of ligand-binding assay suitable for use according to the present invention is the yeast two-hybrid system (Fields and Song, 1989, *Nature*, 340:245-246). The yeast two-hybrid system takes advantage of the properties of the GAL4 protein of the yeast *S. cerevisiae*. The GAL4 protein is a transcriptional activator required for the expression of genes encoding enzymes involving the utilization of galactose. GAL4 protein consists of two separable and functionally essential domains: an N-terminal domain, which binds to specific DNA sequences (UASG); and a C-terminal domain containing acidic regions, which is necessary to activate transcription. The

native GAL4 protein, containing both domains, is a potent activator of transcription when yeast cells are grown on galactose medium. The N-terminal domain binds to DNA in a sequence-specific manner but is unable to activate transcription. The C-terminal domain contains the activating regions but cannot activate transcription because it fails to be localized to UASG. In the two-hybrid system, a system of two hybrid proteins containing parts of GAL4: (1) a GAL4 DNA-binding domain fused to a protein 'X', and (2) a GAL4 activation region fused to a protein 'Y'. If X and Y can form a protein-protein complex and reconstitute proximity of the GAL4 domains, transcription of a gene regulated by UASG occurs. Creation of two hybrid proteins, each containing one of the interacting proteins X and Y, allows the activation region of UASG to be brought to its normal site of action.

The binding assay described in Fodor et al., 1991, *Science*, 251:767-773, which involves testing the binding affinity of test compounds for a plurality of defined polymers synthesized on a solid substrate, may also be useful. Compounds that bind to an HDAC polypeptide, or portions thereof, according to this invention are potentially useful as agents for use in therapeutic compositions.

In another embodiment, sequences encoding an HDAC polypeptide may be synthesized in whole, or in part, using chemical methods well known in the art (See, for example, M.H. Caruthers et al., 1980, *Nucl. Acids Res. Symp. Ser.*, 215-223 and T. Horn, T et al., 1980, *Nucl. Acids Res. Symp. Ser.*, 225-232). Alternatively, an HDAC protein or peptide itself may be produced using chemical methods to synthesize the amino acid sequence of the HDAC polypeptide or peptide, or a fragment or portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (J.Y. Roberge et al., 1995, *Science*, 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (PE Biosystems).

The newly synthesized peptide can be substantially purified by preparative high performance liquid chromatography (e.g., T. Creighton, 1983, *Proteins, Structures and Molecular Principles,* WH Freeman and Co., New

York, N.Y), by reversed-phase high performance liquid chromatography, or other purification methods as are known in the art. The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, *supra*). In addition, the amino acid sequence of an HDAC polypeptide, peptide, or any portion thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

Expression of Human HDAC Proteins

To express a biologically active / functional HDAC polypeptide or peptide, the nucleotide sequences encoding the HDAC polypeptides, or functional equivalents, may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods that are well known to and practiced by those skilled in the art may be used to construct expression vectors containing sequences encoding an HDAC polypeptide or peptide and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described in J. Sambrook et al., 1989, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview, N.Y. and in F.M. Ausubel et al., 1989, *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, N.Y.

A variety of expression vector/host systems may be utilized to contain and express sequences encoding an HDAC polypeptide or peptide. Such expression vector/host systems include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast or fungi transformed with yeast or fungal expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus (CaMV) and tobacco mosaic virus (TMV)), or with bacterial expression vectors

(e.g., Ti or pBR322 plasmids); or animal cell systems. The host cell employed is not limiting to the present invention.

"Control elements" or "regulatory sequences" are those non-translated regions of the vector, e.g., enhancers, promoters, 5' and 3' untranslated regions, which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the BLUESCRIPT phagemid (Stratagene, La Jolla, CA) or PSPORT1 plasmid (Life Technologies), and the like, may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes), or from plant viruses (e.g., viral promoters or leader sequences), may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferred. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding an HDAC polypeptide or peptide, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected, depending upon the use intended for the expressed HDAC product. For example, when large quantities of expressed protein are needed for the induction of antibodies, vectors that direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding an HDAC polypeptide, or peptide, may be ligated into the vector in-frame with sequences for the amino-terminal Met and the subsequent 7 residues of ß-galactosidase, so that a hybrid protein is produced; pIN vectors (See, G. Van Heeke and S.M. Schuster, 1989, *J. Biol. Chem.*, 264:5503-5509); and the like. pGEX vectors (Promega, Madison, WI) may also be used to express foreign

polypeptides, as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can be easily purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. (For reviews, see F.M. Ausubel et al., *supra*, and Grant et al., 1987, *Methods Enzymol.*, 153:516-544).

Should plant expression vectors be desired and used, the expression of sequences encoding an HDAC polypeptide or peptide may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (N. Takamatsu, 1987, *EMBO J.*, 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO, or heat shock promoters, may be used (G. Coruzzi et al., 1984, *EMBO J.*, 3:1671-1680; R. Broglie et al., 1984, *Science*, 224:838-843; and J. Winter et al., 1991, *Results Probl. Cell Differ.* 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (See, for example, S. Hobbs or L.E. Murry, In: McGraw Hill *Yearbook of Science and Technology* (1992) McGraw Hill, New York, N.Y.; pp. 191-196).

An insect system may also be used to express an HDAC polypeptide or peptide. For example, in one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequences encoding an HDAC polypeptide or peptide may be cloned into a non-essential region of the virus such as the polyhedrin gene and placed under control of the polyhedrin promoter. Successful insertion of the HDAC polypeptide or

peptide will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which the HDAC polypeptide or peptide product may be expressed (E.K. Engelhard et al., 1994, *Proc. Nat. Acad. Sci.*, 91:3224-3227).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding an HDAC polypeptide or peptide may be ligated into an adenovirus transcription/translation complex containing the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the HDAC polypeptide or peptide in infected host cells (J. Logan and T. Shenk, 1984, *Proc. Natl. Acad. Sci.*, 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding an HDC polypeptide or peptide. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding an HDAC polypeptide or peptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals, including the ATG initiation codon, should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system that is used, such as those described in the literature (D. Scharf et al., 1994, *Results Probl. Cell Differ.*, 20:125-162).

Moreover, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein

in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells having specific cellular machinery and characteristic mechanisms for such post-translational activities (e.g., COS, CHO, HeLa, MDCK, HEK293, and W138) are available from the American Type Culture Collection (ATCC), American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, and may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express an HDAC protein may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same, or on a separate, vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched cell culture medium before they are switched to selective medium. The purpose of the selectable marker is to confer resistance to selection, and its presence allows the growth and recovery of cells that successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the Herpes Simplex Virus thymidine kinase (HSV TK), (M. Wigler et al., 1977, *Cell*, 11:223-32) and adenine phosphoribosyltransferase (I. Lowy et al., 1980, *Cell*, 22:817-23) genes which can be employed in tk^- or $aprt^-$ cells, respectively. Also, anti-metabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr, which confers resistance to methotrexate (M. Wigler et al., 1980, *Proc. Natl. Acad. Sci.*, 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (F. Colbere-Garapin

et al., 1981, *J. Mol. Biol.*, 150:1-14); and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (S.C. Hartman and R.C. Mulligan, 1988, *Proc. Natl. Acad. Sci.*, 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as the anthocyanins, ß-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, which are widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression that is attributable to a specific vector system (C.A. Rhodes et al., 1995, *Methods Mol. Biol.*, 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the desired gene of interest may need to be confirmed. For example, if an HDAC nucleic acid sequence is inserted within a marker gene sequence, recombinant cells containing sequences encoding the HDAC polypeptide or peptide can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding an HDAC polypeptide or peptide under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates co-expression of the tandem gene.

Alternatively, host cells which contain the nucleic acid sequence encoding an HDAC polypeptide or peptide and which express the HDAC product may be identified by a variety of procedures known to those having skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques, including membrane, solution, or chip based technologies, for the detection and/or quantification of nucleic acid or protein.

Preferably, the HDAC polypeptide or peptide of this invention is substantially purified after expression. HDAC proteins and peptides can be isolated or purified in a variety of ways known to and practiced by those

having skill in the art, depending on what other components may be present in the sample. Standard purification methods include electrophoretic, molecular, immunological and chromatographic techniques, including, but not limited to, ion exchange, hydrophobic affinity and reverse phase HPLC chromatography, and chromatofocusing. For example, an HDAC protein or peptide can be purified using a standard anti-HDAC antibody column. Ultrafiltration and diafiltration techniques, in conjunction with protein concentration, are also useful. For general guidance in suitable purification techniques, see R. Scopes, 1982, *Protein Purification*, Springer-Verlag, NY. As will be understood by the skilled practitioner, the degree of purification necessary will vary depending on the intended use of the HDAC protein or peptide; in some instances, no purification will be necessary.

In addition to recombinant production, fragments of an HDAC polypeptide or peptide may be produced by direct peptide synthesis using solid-phase techniques (J. Merrifield, 1963, *J. Am. Chem. Soc.*, 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using ABI 431A Peptide Synthesizer (PE Biosystems). If desired, various fragments of an HDAC polypeptide can be chemically synthesized separately and then combined using chemical methods to produce the full length molecule.

<u>Detection of Human HDAC Polynucleotide</u>

The presence of polynucleotide sequences encoding an HDAC polypeptide or this invention can be detected by DNA-DNA or DNA-RNA hybridization, or by amplification using probes or portions or fragments of polynucleotides encoding the HDAC polypeptide. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers, based on the sequences encoding a particular HDAC polypeptide or peptide, to detect transformants containing DNA or RNA encoding an HDAC polypeptide or peptide.

A wide variety of labels and conjugation techniques are known and employed by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR

probes for detecting sequences related to polynucleotides encoding an HDAC polypeptide or peptide include oligo-labeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding an HDAC polypeptide, or any portions or fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase, such as T7, T3, or SP(6) and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits (e.g., Amersham Pharmacia Biotech, Promega and U.S. Biochemical Corp.).

Suitable reporter molecules or labels which may be used include radionucleotides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like. Non-limiting examples of labels include radioisotopes, such as ^{3}H, ^{14}C, and ^{32}P, and non-radioactive molecules, such as digoxigenin. In addition, nucleic acid molecules may be modified using known techniques, for example, using RNA or DNA analogs, phosphorylation, dephosphorylation, methylation, or demethylation.

<u>Human HDAC Polypeptides – Production, Detection, Isolation</u>

Host cells transformed with nucleotide sequences encoding an HDAC protein or peptide, or fragments thereof, may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those having skill in the art, expression vectors containing polynucleotides which encode an HDAC protein or peptide may be designed to contain signal sequences that direct secretion of the HDAC protein or peptide through a prokaryotic or eukaryotic cell membrane.

Other constructions may be used to join nucleic acid sequences encoding an HDAC protein or peptide to a nucleotide sequence encoding a polypeptide domain that will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating

peptides such as histidine-tryptophan modules that allow purification on immobilized metals; protein A domains that allow purification on immobilized immunoglobulin; and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and the HDAC protein or peptide may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing HDAC-encoding sequence and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMAC (immobilized metal ion affinity chromatography) as described by J. Porath et al., 1992, *Prot. Exp. Purif.*, 3:263-281, while the enterokinase cleavage site provides a means for purifying from the fusion protein. For a discussion of suitable vectors for fusion protein production, see D.J. Kroll et al., 1993; *DNA Cell Biol.*, 12:441-453.

Human artificial chromosomes (HACs) may be used to deliver larger fragments of DNA than can be contained and expressed in a plasmid vector. HACs are linear microchromosomes which may contain DNA sequences of 10K to 10M in size, and contain all of the elements that are required for stable mitotic chromosome segregation and maintenance (See, J.J. Harrington et al., 1997, *Nature Genet.*, 15:345-355). HACs of 6 to 10M are constructed and delivered via conventional delivery methods (e.g., liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

A variety of protocols for detecting and measuring the expression of an HDAC polypeptide using either polyclonal or monoclonal antibodies specific for the protein are known and practiced in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive with two non-interfering epitopes on the HDAC polypeptide is preferred, but a competitive binding assay may also be employed. These and other assays are described in the art as represented by the publication of R. Hampton et al., 1990; *Serological*

Methods, a Laboratory Manual, APS Press, St Paul, MN and D.E. Maddox et al., 1983; *J. Exp. Med.*, 158:1211-1216).

For use with these assays, amino acid sequences (e.g., polypeptides, peptides, antibodies, or antibody fragments) may be attached to a label capable of providing a detectable signal, either directly or indirectly, including, but not limited to, radioisotope, fluorescent, and enzyme labels. Fluorescent labels include, for example, Cy3, Cy5, Alexa, BODIPY, fluorescein (e.g., FluorX, DTAF, and FITC), rhodamine (e.g., TRITC), auramine, Texas Red, AMCA blue, and Lucifer Yellow. Preferred isotope labels include ^{3}H, ^{14}C, ^{32}P, ^{35}S, ^{36}Cl, ^{51}Cr, ^{57}Co, ^{58}Co, ^{59}Fe, ^{90}Y, ^{125}I, ^{131}I, and ^{186}Re. Preferred enzyme labels include peroxidase, β-glucuronidase, β-D-glucosidase, β-D-galactosidase, urease, glucose oxidase plus peroxidase, and alkaline phosphatase (see, e.g., U.S. Pat. Nos. 3,654,090; 3,850,752 and 4,016,043). Enzymes can be conjugated by reaction with bridging molecules such as carbodiimides, diisocyanates, glutaraldehyde, and the like. Enzyme labels can be detected visually, or measured by calorimetric, spectrophotometric, fluorospectrophotometric, amperometric, or gasometric techniques. Other labeling systems, such as avidin/biotin, Tyramide Signal Amplification (TSA™), are known in the art, and are commercially available (see, e.g., ABC kit, Vector Laboratories, Inc., Burlingame, CA; NEN® Life Science Products, Inc., Boston, MA).

A compound that interacts with a histone deacetylase according to the present invention may be one that is a substrate for the enzyme, one that binds the enzyme at its active site, or one that otherwise acts to alter enzyme activity by binding to an alternate site. A substrate may be acetylated histones, or a labeled acetylated peptide fragment derived therefrom, such as AcGly-Ala-Lys,(.epsilon.-Ac)-Arg-His-Arg-Lys,(.epsilon.-Ac)-ValNH_2, or other synthetic or naturally occurring substrates. Examples of compounds that bind to histone deacetylase are known inhibitors such as n-butyrate, trichostatin, trapoxin and SAHA (S. Swendeman et al., 1999, *Cancer Res.*, 59(17):4392-4399). The compound that interacts with a histone deacetylase is preferably

labeled to allow easy quantification of the level of interaction between the compound and the enzyme. A preferred radiolabel is tritium.

The test compound (i.e., test agent) may be a synthetic compound, a purified preparation, crude preparation, or an initial extract of a natural product obtained from plant, microorganism or animal sources.

One aspect of the present method is based on test compound- induced inhibition of histone deacetylase activity. The enzyme inhibition assay involves adding histone deacetylase or an extract containing histone deacetylase to mixtures of an enzyme substrate and the test compound, both of which are present in known concentrations. The amount of the enzyme is chosen such that approximately 20% of the substrate is consumed during the assay. The assay is carried out with the test compound at a series of different dilution levels. After a period of incubation, the labeled portion of the substrate released by enzymatic action is separated and counted. The assay is generally carried out in parallel with a negative control (i.e., no test compound) and a positive control (i.e., containing a known enzyme inhibitor instead of a test compound). The concentration of the test compound at which 50% of the enzyme activity is inhibited (IC_{50}) is determined using art recognized method.

Although enzyme inhibition is the most direct measure of the inhibitory activity of the test compound, results obtained from a competitive binding assay in which the test compound competes with a known inhibitor for binding to the enzyme active site correlate well with the results obtained from enzyme inhibition assay described above. The binding assay represents a more convenient way to assess enzyme inhibition, because it allows the use of a crude extract containing histone deacetylase rather than partially purified enzyme. The use of a crude extract may not always be suitable in the enzyme inhibition assay because other enzymes present in the extract may act on the histone deacetylase substrate.

The competition binding assay is carried out by adding a histone deacetylase, or an extract containing histone deacetylase activity, to a mixture of the test compound and a labeled inhibitor, both of which are present in the

mixture in known concentrations. After incubation, the enzyme-inhibitor complex is separated from the unbound labeled inhibitors and unlabeled test compound, and counted. The concentration of the test compound required to inhibit 50% of the binding of the labeled inhibitor to the histone deacetylase (IC_{50}) is calculated.

In one method suitable for this invention, the IC_{50} of test compounds against host histone deacetylase is determined using either the enzyme inhibition assay or the binding assay as described above, to identify those compounds that have selectivity for a particular type of histone deacetylase over that of a host.

Anti-Human HDAC Antibodies and Uses Thereof

Antagonists or inhibitors of the HDAC polypeptides of the present invention may be produced using methods that are generally known in the art. In particular, purified HDAC polypeptides or peptides, or fragments thereof, can be used to produce antibodies, or to screen libraries of pharmaceutical agents or other compounds, particularly, small molecules, to identify those which specifically bind to the novel HDACs of this invention.

Antibodies specific for an HDAC polypeptide, or immunogenic peptide fragments thereof, can be generated using methods that have long been known and conventionally practiced in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments, and fragments produced by an Fab expression library. Neutralizing antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, sheep, rats, mice, humans, and others, can be immunized by injection with HDAC polypeptide, or any peptide fragment or oligopeptide thereof, which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase the immunological response. Nonlimiting examples of suitable adjuvants include Freund's (incomplete), mineral gels such as aluminum hydroxide or silica, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and

dinitrophenol. Adjuvants typically used in humans include BCG (bacilli Calmette Guérin) and *Corynebacterium parvumn*.

Preferably, the peptides, fragments, or oligopeptides used to induce antibodies to HDAC polypeptides (i.e., immunogens) have an amino acid sequence having at least five amino acids, and more preferably, at least 7-10 amino acids. It is also preferable that the immunogens are identical to a portion of the amino acid sequence of the natural protein; they may also contain the entire amino acid sequence of a small, naturally occurring molecule. The peptides, fragments or oligopeptides may comprise a single epitope or antigenic determinant or multiple epitopes. Short stretches of HDAC amino acids may be fused with those of another protein, such as KLH, and antibodies are produced against the chimeric molecule.

Monoclonal antibodies to HDAC polypeptides, or immunogenic fragments thereof, may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (G. Kohler et al., 1975, *Nature*, 256:495-497; D. Kozbor et al., 1985, *J. Immunol. Methods*, 81:31-42; R.J. Cote et al., 1983, *Proc. Natl. Acad. Sci. USA*, 80:2026-2030; and S.P. Cole et al., 1984, *Mol. Cell Biol.*, 62:109-120). The production of monoclonal antibodies is well known and routinely used in the art.

In addition, techniques developed for the production of "chimeric antibodies," the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (S.L. Morrison et al., 1984, *Proc. Natl. Acad. Sci. USA*, 81:6851-6855; M.S. Neuberger et al., 1984, *Nature*, 312:604-608; and S. Takeda et al., 1985, *Nature*, 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce HDAC polypeptide- or peptide-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (D.R. Burton, 1991, *Proc. Natl. Acad. Sci. USA*,

88:11120-3). Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (R. Orlandi et al., 1989, *Proc. Natl. Acad. Sci. USA*, 86:3833-3837 and G. Winter et al., 1991, *Nature*, 349:293-299).

Antibody fragments that contain specific binding sites for an HDAC polypeptide or peptide may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments which can be produced by pepsin digestion of the antibody molecule and Fab fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (W.D. Huse et al., 1989, *Science*, 254.1275-1281).

Various immunoassays can be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve measuring the formation of complexes between an HDAC polypeptide and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive with two non-interfering HDAC epitopes is preferred, but a competitive binding assay may also be employed (Maddox, *supra*).

Antibodies which specifically bind HDAC epitopes can also be used in immunohistochemical staining of tissue samples to evaluate the abundance and pattern of expression of each of the provided HDAC polypeptides. Anti-HDAC antibodies can be used diagnostically in immuno-precipitation and immunoblotting techniques to detect and evaluate HDAC protein levels in tissue as part of a clinical testing procedure. For instance, such measurements can be useful in predictive evaluations of the onset or progression of proliferative or differentiation disorders. Similarly, the ability to monitor HDAC protein levels in an individual can allow the determination of the efficacy of a given treatment regimen for an individual afflicted with such a

disorder. The level of HDAC polypeptide may be measured from cells in a bodily fluid, such as in samples of cerebral spinal fluid or amniotic fluid, or can be measured in tissue, such as produced by biopsy. Diagnostic assays using anti-HDAC antibodies can include, for example, immunoassays designed to aid in early diagnosis of a disorder, particularly ones that are manifest at birth. Diagnostic assays using anti-HDAC polypeptide antibodies can also include immunoassays designed to aid in early diagnosis and phenotyping of neoplastic or hyperplastic disorders.

Another application of anti-HDAC antibodies according to the present invention is in the immunological screening of cDNA libraries constructed in expression vectors such as λgt11, λgt 18-23, λZAP, and λORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For example, λgt11 will produce fusion proteins whose amino termini contain 13-galactosidase amino acid sequences and whose carboxy termini contain a foreign polypeptide. Antigenic epitopes of an HDAC protein, e.g. other orthologs of a particular HDAC protein or other paralogs from the same species, can then be detected with antibodies by, for example, reacting nitrocellulose filters lifted from infected plates with anti-HDAC antibodies. Positive phage detected by this assay can then be isolated from the infected plate. Thus, the presence of HDAC homologs can be detected and cloned from other animals, as can alternative isoforms (including splice variants) from humans.

<u>Therapeutics/Treatments/Methods of Use Involving HDACs</u>

In an embodiment of the present invention, the polynucleotide encoding an HDAC polypeptide or peptide, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, antisense to the polynucleotide encoding a novel HDAC polypeptide may be used in situations in which it would be desirable to block the transcription of HDAC mRNA. In particular, cells may be transformed or transfected with sequences complementary to polynucleotides encoding an HDAC polypeptide. Thus, complementary molecules may be used to modulate human HDAC polynucleotide and polypeptide activity, or to achieve regulation of gene

function. Such technology is now well known in the art, and sense or antisense oligomers or oligonucleotides, or larger fragments, can be designed from various locations along the coding or control regions of polynucleotide sequences encoding the HDAC polypeptides. For antisense therapeutics, the oligonucleotides in accordance with this invention preferably comprise at least 3 to 50 nucleotides of a sequence complementary to SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, or SEQ ID NO:96. It is more preferred that such oligonucleotides and analogs comprise at least 8 to 25 nucleotides, and still more preferred to comprise at least 12 to 20 nucleotides of this sequence.

Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct recombinant vectors which will express nucleic acid sequences that are complementary to the nucleic acid sequences encoding the novel HDAC polypeptides and peptides of the present invention. These techniques are described both in J. Sambrook et al., *supra* and in F.M. Ausubel et al., *supra*.

A preferred approach for *in vivo* introduction of nucleic acid into a cell is by use of a viral vector containing nucleic acid, e.g. a cDNA encoding the particular HDAC polypeptide desired. Infection of cells with a viral vector has the advantage that a large proportion of the targeted cells can receive the nucleic acid. In addition, molecules encoded within the viral vector, e.g., by a cDNA contained in the viral vector, are expressed efficiently in cells that have taken up viral vector nucleic acid. As mentioned, retrovirus vectors, adenovirus vectors and adeno-associated virus vectors are exemplary recombinant gene delivery system for the transfer of exogenous genes *in vivo*, particularly into humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host.

In addition to the above-illustrated viral transfer methods, non-viral methods can also be employed to yield expression of an HDAC polypeptide in

the cells and/or tissue of an animal. Most non-viral methods of gene transfer rely on normal mechanisms used by mammalian cells for the uptake and intracellular transport of macromolecules. In preferred embodiments, non-viral gene delivery systems rely on endocytic pathways for the uptake of the novel HDAC polypeptide-encoding gene by the targeted cell. Exemplary gene delivery systems of this type include liposomal derived systems, poly-lysine conjugates, and artificial viral envelopes.

In clinical settings, the gene delivery systems for a therapeutic HDAC gene can be introduced into a patient by any of a number of methods, each of which is familiar in the art. For instance, a pharmaceutical preparation of the gene delivery system can be introduced systematically, e.g., by intravenous injection, and specific transduction of the protein in the target cells occurs predominantly from the specificity of transfection provided by the gene delivery vehicle, cell-type or tissue-type expression due to the transcriptional regulatory sequences controlling expression of the receptor gene, or a combination thereof.

In other aspects. the initial delivery of a recombinant HDAC gene is more limited, for example, with introduction into an animal being quite localized. For instance, the gene delivery vehicle can be introduced by catheter (see, U.S. Patent No. 5,328,470) or by stereotactic injection (e.g., Chen et al., 1994, *Proc. Natl. Acad. Sci. USA*, 91:3054-3057). An HDAC nucleic acid sequence (gene), e.g., sequences represented by SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, and/or SEQ ID NO:96, or a fragment thereof, can be delivered in a gene therapy construct by electroporation using techniques described, for example, by Dev et al. (1994, *Cancer Treat. Rev.*, 20:105-115).

The gene encoding an HDAC polypeptide can be turned off by transforming a cell or tissue with an expression vector that expresses high levels of an HDAC polypeptide-encoding polynucleotide, or a fragment thereof. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are

disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and even longer if appropriate replication elements are designed to be part of the vector system.

Modifications of gene expression can be obtained by designing antisense molecules or complementary nucleic acid sequences (DNA, RNA, or PNA), to the control, 5', or regulatory regions of the genes encoding the novel HDAC polypeptides, (e.g., signal sequence, promoters, enhancers, and introns). Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferable. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described (See, for example, J.E. Gee et al., 1994, In: B.E. Huber and B.I. Carr, *Molecular and Immunologic Approaches,* Futura Publishing Co., Mt. Kisco, NY). The antisense molecule or complementary sequence may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, i.e., enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Suitable examples include engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding the HDAC polypeptides.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be

evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes according to the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. Such methods include techniques for chemically synthesizing oligonucleotides, for example, solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding the human HDACs of the present invention. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP. Alternatively, the cDNA constructs that constitutively or inducibly synthesize complementary HDAC RNA can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl (rather than phosphodiesterase linkages) within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and are equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection and by liposome injections may be achieved using methods that are well known in the art.

In another embodiment of the present invention, an expression vector containing the complement of the polynucleotide encoding an HDAC polypeptide, or an antisense HDAC oligonucleotide, may be administered to

an individual to treat or prevent a disease or disorder associated with uncontrolled or neoplastic cell growth, hyperactivity or stimulation, for example. A variety of specialized oligonucleotide delivery techniques may be employed, for example, encapsulation in unilamellar liposomes and reconstituted Sendai virus envelopes for RNA and DNA delivery (Arad et al., 1986, *Biochem. Biophys. Acta.*, 859:88-94).

In another embodiment, the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the present invention can be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

Any of the therapeutic methods described above may be applied to any individual in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

Another aspect of the present invention involves a method for modulating one or more of growth, differentiation, or survival of a mammalian cell by modulating HDAC bioactivity, e.g., by inhibiting the deacetylase activity of HDAC proteins, or disrupting certain protein-protein interactions. In general, whether carried out *in vivo*, *in vitro*, *ex vivo*, or *in situ*, the method comprises treating a cell with an effective amount of an HDAC therapeutic so as to alter, relative to an effect in the absence of treatment, one or more of (i) rate of growth or proliferation, (ii) differentiation, or (iii) survival of the cell. Accordingly, the method can be carried out with HDAC therapeutics, such as peptide and peptidomimetics, or other molecules identified in the drug screening methods as described herein which antagonize the effects of a naturally-occurring HDAC protein on a cell.

Other HDAC therapeutics include antisense constructs for inhibiting expression of HDAC proteins, and dominant negative mutants of HDAC proteins which competitively inhibit protein-substrate and/or protein- protein interactions upstream and downstream of the wild-type HDAC protein. In an exemplary embodiment, an antisense method is used to treat tumor cells by antagonizing HDAC activity and blocking cell cycle progression. The method includes, but is not limited to, the treatment of testicular cells, so as modulate spermatogenesis; the modulation of osteogenesis or chondrogenesis, comprising the treatment of osteogenic cells or chondrogenic cell, respectively, with an HDAC polypeptide. In addition, HDAC polypeptides can be used to modulate the differentiation of progenitor cells, e.g., the method can be used to cause differentiation of hematopoietic cells, neuronal cells, or other stem/progenitor cell populations, to maintain a cell in a differentiated state, and/or to enhance the survival of a differentiated cell, e.g., to prevent apoptosis or other forms of cell death.

The present method is applicable, for example, to cell culture techniques, such as in the culturing of hematopoietic cells and other cells whose survival or differentiation state is dependent on HDAC function. Moreover, HDAC agonists and antagonists can be used for therapeutic intervention, such as to enhance survival and maintenance of cells, as well as to influence organogenic pathways, such as tissue patterning and other differentiation processes. As an example, such a method is practiced for modulating, in an animal, cell growth, cell differentiation or cell survival, and comprises administering a therapeutically effective amount of an HDAC polypeptide to alter, relative the absence of HDAC treatment, one or more of (i) rate of cell growth or proliferation, (ii) cell differentiation, and/or (iii) cell survival of one or more cell types in an animal.

In another of its aspects the present invention provides a method of determining if a subject, e.g., a human patient, is at risk for a disorder characterized by unwanted cell proliferation or aberrant control of differentiation. The method includes detecting, in a tissue of the subject, the presence or the absence of a genetic lesion characterized by at least one of

(i) a mutation of a gene encoding an HDAC protein, e.g. represented in one of SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, or SEQ ID NO:96, or a homolog thereof, or (ii) the mis-expression of an HDAC gene. More specifically, detecting the genetic lesion includes ascertaining the existence of at least one of a deletion of one or more nucleotides from an HDAC gene; an addition of one or more nucleotides to the gene, a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene; an alteration in the level of a messenger RNA transcript of the gene; the presence of a non-wild type splicing pattern of an mRNA transcript of the gene; or a non-wild type level of the protein.

For example, detecting a genetic lesion can include (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence of an HDAC gene, e.g., a nucleic acid represented in one of SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, or SEQ ID NO:96, or naturally occurring mutants thereof, or 5' or 3' flanking sequences naturally associated with the HDAC gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and (iii) detecting, by hybridization of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion; e.g., wherein detecting the lesion comprises utilizing the probe/primer to determine the nucleotide sequence of the HDAC gene and, optionally, of the flanking nucleic acid sequences. For instance, the probe/primer can be employed in a polymerase *chain reaction* (PCR) or in a *ligation chain reaction* (LCR). In alternative embodiments, the level of an HDAC protein is detected in an immunoassay using an antibody that is specifically immunoreactive with the HDAC protein.

<u>Methods And Therapeutic Uses Related To Cell Modulation</u>

Another aspect of the present invention relates to a method of inducing and/or maintaining a differentiated state, enhancing survival, and/or inhibiting (or alternatively, potentiating) the proliferation of a cell, by contacting cells with an agent that modulates HDAC-dependent transcription. In view of the apparently broad involvement of HDAC proteins in the control of chromatin

structure and, in turn, transcription and replication, the present invention contemplates a method for generating and/or maintaining an array of different tissue both *in vitro* and *in vivo*. An "HDAC therapeutic," whether inhibitory or potentiating with respect to modulating histone deacetylation, can be, as appropriate, any of the preparations described herein, including isolated polypeptides, gene therapy constructs, antisense molecules, peptidomimetics, or agents identified in the drug and bioactive screening assays and methods described herein.

As an aspect of the present invention, the HDAC modulatory (i.e., inhibitory or stimulatory) compounds are likely to play an important role in effecting cellular proliferation. There are a wide variety of pathological cell proliferative conditions for which HDAC therapeutic agents of the present invention may be used in treatment. For instance, such agents can provide therapeutic benefits in the inhibition of an anomalous cell proliferation. Nonlimiting examples of diseases and conditions that may benefit from such methods include various cancers and leukemias, psoriasis, bone diseases, fibroproliferative disorders, e.g., those involving connective tissues, atherosclerosis and other smooth muscle proliferative disorders, as well as chronic inflammation.

Non-limiting cancer types include carcinoma (e.g., adenocarcinoma), sarcoma, myeloma, leukemia, and lymphoma, and mixed types of cancers, such as adenosquamous carcinoma, mixed mesodermal tumor, carcinosarcoma, and teratocarcinoma. Representative cancers include, but are not limited to, bladder cancer, lung cancer, breast cancer, colon cancer, rectal cancer, endometrial cancer, ovarian cancer, head and neck cancer, prostate cancer, and melanoma. Specifically included are AIDS-related cancers (e.g., Kaposi's Sarcoma, AIDS-related lymphoma), bone cancers (e.g., osteosarcoma, malignant fibrous histiocytoma of bone, Ewing's Sarcoma, and related cancers), and hematologic/blood cancers (e.g., adult acute lymphoblastic leukemia, childhood acute lymphoblastic leukemia, adult acute myeloid leukemia, childhood acute myeloid leukemia, chronic lymphocytic leukemia, chronic myelogenous leukemia, hairy cell leukemia,

cutaneous T-cell lymphoma, adult Hodgkin's disease, childhood Hodgkin's disease, Hodgkin's disease during pregnancy, mycosis fungoides, adult non-Hodgkin's lymphoma, childhood non-Hodgkin's lymphoma, non-Hodgkin's lymphoma during pregnancy, primary central nervous system lymphoma, Sezary syndrome, cutaneous T-cell lymphoma, Waldenström's macroglobulinemia, multiple myeloma/plasma cell neoplasm, myelodysplastic syndrome, and myeloproliferative disorders).

Also included are brain cancers (e.g., adult brain tumor, childhood brain stem glioma, childhood cerebellar astrocytoma, childhood cerebral astrocytoma, childhood ependymoma, childhood medulloblastoma, supratentorial primitive neuroectodermal and pineal, and childhood visual pathway and hypothalamic glioma), digestive/gastrointestinal cancers (e.g., anal cancer, extrahepatic bile duct cancer, gastrointestinal carcinoid tumor, colon cancer, esophageal cancer, gallbladder cancer, adult primary liver cancer, childhood liver cancer, pancreatic cancer, rectal cancer, small intestine cancer, and gastric cancer), musculoskeletal cancers (e.g., childhood rhabdomyosarcoma, adult soft tissue sarcoma, childhood soft tissue sarcoma, and uterine sarcoma), and endocrine cancers (e.g., adrenocortical carcinoma, gastrointestinal carcinoid tumor, islet cell carcinoma (endocrine pancreas), parathyroid cancer, pheochromocytoma, pituitary tumor, and thyroid cancer).

Further included are neurologic cancers (e.g., neuroblastoma, pituitary tumor, and primary central nervous system lymphoma), eye cancers (e.g., intraocular melanoma and retinoblastoma), genitourinary cancers (e.g., bladder cancer, kidney (renal cell) cancer, penile cancer, transitional cell renal pelvis and ureter cancer, testicular cancer, urethral cancer, Wilms' tumor and other childhood kidney tumors), respiratory/thoracic cancers (e.g., non-small cell lung cancer, small cell lung cancer, malignant mesothelioma, and malignant thymoma), germ cell cancers (e.g., childhood extracranial germ cell tumor and extragonadal germ cell tumor), skin cancers (e.g., melanoma, and merkel cell carcinoma), gynecologic cancers (e.g., cervical cancer, endometrial cancer, gestational trophoblastic tumor, ovarian epithelial cancer,

ovarian germ cell tumor, ovarian low malignant potential tumor, uterine sarcoma, vaginal cancer, and vulvar cancer), and unknown primary cancers.

In certain aspects of the inventions, the disclosed HDAC inhibitors, antisense molecules, anti-HDAC antibodies, or antibody fragments can be used as treatments for breast or prostate cancers. In particular, HDAC9c inhibitors, HDAC9c antisense molecules, anti-HDAC9c antibodies, or fragments thereof, can be used. Specific breast cancers include, but are not limited to, non-invasive cancers, such as ductal carcinoma *in situ* (DCIS), intraductal carcinoma lobular carcinoma *in situ* (LCIS), papillary carcinoma, and comedocarcinoma, or invasive cancers, such as adenocarcinomas, or carcinomas, e.g., infiltrating ductal carcinoma, infiltrating lobular carcinoma, infiltrating ductal and lobular carcinoma, medullary carcinoma, mucinous (colloid) carcinoma, comedocarcinoma, Paget's Disease, papillary carcinoma, tubular carcinoma, and inflammatory carcinoma. Specific prostate cancers may include adenocarcinomas and sarcomas, or pre-cancerous conditions, such as prostate intraepithelial neoplasia (PIN).

In addition to proliferative disorders, the present invention envisions the use of HDAC therapeutics for the treatment of differentiation disorders resulting from, for example, de-differentiation of tissue which may (optionally) be accompanied by abortive reentry into mitosis, e.g. apoptosis. Such degenerative disorders include chronic neurodegenerative diseases of the nervous system, including Alzheimer's disease, Parkinson's disease, Huntington's chorea, amylotrophic lateral sclerosis (ALS) and the like, as well as spinocerebellar degenerations. Other differentiation disorders include, for example, disorders associated with connective tissue, such as can occur due to de-differentiation of chondrocytes or osteocytes, as well as vascular disorders which involve de-differentiation of endothelial tissue and smooth muscle cells, gastric ulcers characterized by degenerative changes in glandular cells, and renal conditions marked by failure to differentiate, e.g. Wilm's tumors.

It will also be recognized that, by transient use of modulators of HDAC activities, *in vivo* reformation of tissue can be accomplished, for example, in

the development and maintenance of organs. By controlling the proliferative and differentiation potential for different cell types, HDAC therapeutics can be used to re-form injured tissue, or to improve grafting and morphology of transplanted tissue. As an example, HDAC antagonists and agonists can be employed in a differential manner to regulate different stages of organ repair after physical, chemical or pathological insult or injury. Such regimens can be utilized, for example, in the repair of cartilage, increasing bone density, liver repair subsequent to a partial hepatectomy, or to promote regeneration of lung tissue in the treatment of emphysema.
The present method is also applicable to cell culture techniques.

More specifically, HDAC therapeutics can be used to induce differentiation of uncommitted progenitor cells, thus giving rise to a committed progenitor cell, or causing further restriction of the developmental fate of a committed progenitor cell toward becoming a terminally differentiated cell. As an example, methods involving HDAC therapeutics can be used *in vitro*, *ex vivo*, or *in vivo* to induce and/or to maintain the differentiation of hematopoietic cells into erythrocytes and other cells of the hematopoietic cell lineage. Illustratively, the effect of erythropoietin (EPO) on the growth of EPO-responsive erythroid precursor cells is increased to influence their differentiation into red blood cells. Also, as an example, the amount of EPO, or other differentiating agent, that is required for growth and/or differentiation is reduced based on the administration of an inhibitor of histone deacetylation. (PCT/US92/07737).

Accordingly, HDAC therapeutics as described, particularly those that antagonize HDAC deacetylase activity, can be administered alone or in conjunction with EPO, for example, in a suitable carrier, to vertebrates to promote erythropoiesis. Alternatively, *ex vivo* cell treatments are suitable. Similar types of treatments can be used for a variety of disease states, including use in individuals who require bone marrow transplants (e.g., patients with aplastic anemia, acute leukemias, recurrent lymphomas, or solid tumors). As an example, prior to receiving a bone marrow transplant, a recipient is prepared by ablating or removing endogenous hematopoietic stem

cells. Such treatment is typically performed by total body irradiation, or by delivery of a high dose of an alkylating agent or other chemotherapeutic cytotoxic agent (Anklesaria et al., 1987, *Proc. Natl. Acad. Sci. USA*), 84:7681-7685). Following the preparation of the recipient, donor bone marrow cells are injected intravenously. Optionally, HDAC therapeutics could be contacted with the cells *ex vivo* or administered to the subject with the re-implanted cells.

In addition, there may be cell-type specific HDAC proteins, and/or some cell types may be more sensitive to the modulation of HDAC deacetylase activities. Even within a cell type, the stage of differentiation or position in the cell cycle could influence a cell's response to a modulatory HDAC therapeutic agent. Accordingly, the present invention contemplates the use of agents that modulate histone deacetylase activity to specifically inhibit or activate certain cell types. As an illustrative example, T cell proliferation could be preferentially inhibited so as to induce tolerance by a procedure similar to that used to induce tolerance using sodium butyrate (see, for example, PCT/US93/03045). Accordingly, HDAC therapeutics may be used to induce antigen specific tolerance in any situation in which it is desirable to induce tolerance, such as autoimmune diseases, in allogeneic or xenogeneic transplant recipients, or in graft versus host (GVH) reactions. Tolerance is typically induced by presenting the tolerizing compound (e.g., an HDAC inhibitor compound) substantially concurrently with the antigen, i.e., within a time period that is reasonably close to that in which the antigen is administered. Preferably, the HDAC therapeutic is administered after presentation of the antigen, so that the cumulative effect will occur after the particular repertoire of T_H cells begins to undergo clonal expansion. Additionally, the present invention contemplates the application of HDAC therapeutics for modulating morphogenic signals involved in organogenic pathways. Thus, it is apparent that compositions comprising HDAC therapeutics can be employed for both cell culture and therapeutic methods involving the generation and maintenance of tissue.

In a further aspect, HDAC therapeutics are useful in increasing the amount of protein produced by a cell, including a recombinant cell. Suitable cells may comprise any primary cell isolated from any animal, cultured cells, immortalized cells, transfected or transformed cells, and established cell lines. Animal cells preferably will include cells which intrinsically have an ability to produce a desired protein; cells which are induced to have an ability to produce a desired protein, for example, by stimulation with a cytokine such as an interferon or an interleukin; genetically engineered cells into which a gene encoding a desired protein is introduced. The protein produced by the process can include peptides or proteins, including peptide-hormone or proteinaceous hormones such as any useful hormone, cytokine, interleukin, or protein which it may be desirable to be produced in purified form and/or in large quantity.

In specific aspects, the HDAC inhibitors, antisense molecules, anti-HDAC antibodies, or antibody fragments of the invention can be used in combination with other HDAC inhibitory agents, e.g., trichostatin A (D.M. Vigushin et al., 2001, *Clin. Cancer Res.* 7(4):971-6); trapoxin A (Itazaki et al., 1990, *J. Antibiot.* 43:1524-1532), MS-275 (T. Suzuki et al., 1999, *J. Med. Chem.* 42(15):3001-3), CHAPs (Y. Komatsu et al., 2001, *Cancer Res.* 61(11):4459-66), CI-994 (see, e.g., P.M. LoRusso et al., 1996, *New Drugs* 14(4):349-56), SAHA (V.M. Richon et al., 2001, *Blood Cells Mol. Dis.* 27(1):260-4), depsipeptide (FR901228; FK228; V. Sandor et al., 2002, *Clin. Cancer Res.* 8(3):718-28), CBHA (D.C. Coffey et al., 2001, *Cancer Res.* 61(9):3591-4), pyroxamide, (L.M. Butler et al, 2001, *Clin. Cancer Res.* 7(4):962-70), CHAP31 (Y. Komatsu et al., 2001, *Cancer Res.* 61(11):4459-66), HC-toxin (Liesch et al., 1982, *Tetrahedron* 38:45-48), chlamydocin (Closse et al., 1974, *Helv. Chim. Acta* 57:533-545), Cly-2 (Hirota et al., 1973, *Agri. Biol. Chem.* 37:955-56), WF-3161 (Umehana et al., 1983, *J. Antibiot.* 36, 478-483; M. Kawai et al., 1986, *J. Med. Chem.* 29(11):2409-11), Tan-1746 (Japanese Patent No. 7196686 to Takeda Yakuhin Kogyo KK), apicidin (S.H. Kwon et al., 2002, *J. Biol. Chem.* 277(3):2073-80), and analogs thereof.

Screening Methods

The novel HDAC proteins, peptides and nucleic acids can be used in screening assays to identify candidate bioactive agents or drugs that modulate HDAC bioactivity, preferably HDAC inhibitors, for potential use to treat neoplastic disorders, for example, to kill cancer cells and tumor cells exhibiting uncontrolled cell growth for numerous reasons, e.g., the lack of a suppressor molecule such as p53. In addition, HDAC proteins and encoding nucleic acids, as well as the bioactive agents that modulate HDAC activity or function, can be used as effectors in methods to regulate cell growth, e.g., to kill neoplastic cells.

The HDAC polynucleotides and polypeptides can also be modulated by interactive molecules. By "modulate" herein is meant that the bioactivity of HDAC is altered, i.e., either increased or decreased. In a preferred embodiment, HDAC function is inhibited. The HDACs can be used as targets to screen for inhibitors of HDAC, e.g., naturally-occurring HDAC, function, bioactivity, or expression in neoplastic cells and/or uncontrolled cell growth. Examples of HDAC biological activity include the ability to modulate the proliferation of cells. For example, inhibiting histone deacetylation causes cells to arrest in the G1 and G2 phases of the cell cycle. The biochemical activity associated with the novel HDAC proteins of the present invention are also characterized in terms of binding to and (optionally) catalyzing the deacetylation of an acetylated histone. Another biochemical property of certain HDAC proteins involves binding to other cellular proteins, such as RbAp48 (Qian et al., 1993, *Nature*, 364:648), or Sin3A. (see, e.g., WO 97/35990)

Generally, in performing screening methods, HDAC polypeptide or peptide can be non-diffusably bound to an insoluble support having isolated sample receiving areas (e.g. a microtiter plate, an array, etc.). The criteria for suitable insoluble supports are that they can be made of any composition to which polypeptides can be bound; they are readily separated from soluble material; and they are otherwise compatible with the overall method of screening. The surface of such supports may be solid or porous and of any

convenient size or shape. Examples of suitable insoluble supports include microtiter plates, arrays, membranes and beads. These are typically made of glass, plastic (e.g., polystyrene), polysaccharides, nylon or nitrocellulose. Microtiter plates and arrays are especially convenient, because a large number of assays can be carried out simultaneously, using small amounts of reagents and samples. The particular manner of binding the polypeptide is not crucial, so long as it is compatible with the reagents and overall methods of the invention, maintains the activity of the peptide and is nondiffusable.

Preferred methods of binding include the use of antibodies (which should not hinder the binding of HDACs to associated proteins), direct binding to "sticky" or ionic supports, chemical crosslinking, etc. Following binding of the polypeptide, excess unbound material is removed by washing. The sample receiving areas may then be blocked as needed through incubation with bovine serum albumin (BSA), casein or other innocuous/nonreactive protein.

A candidate bioactive agent is added to the assay. Novel binding agents include specific antibodies, non-natural binding agents identified in screens of chemical libraries, peptide analogs, etc. Of particular interest are screening assays for agents that have a low toxicity for human cells. A wide variety of assays may be used for this purpose, including labeled *in vitro* protein-protein binding assays, electrophoretic mobility shift assays, immunoassays for protein binding, and the like. The term "agent" as used herein describes any molecule, e.g., protein, oligopeptide, small organic molecule, polysaccharide, polynucleotide, etc., having the capability of directly or indirectly altering the activity or function of HDAC polypeptides. Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, i.e., at zero concentration, or below the level of detection.

Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 100 and less than about 10,000

daltons, preferably, less than about 2000 to 5000 daltons, as a nonlimiting example. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. In addition, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification to produce structural analogs.

The determination of the binding of the candidate biomolecule or agent to an HDAC polypeptide may be accomplished in a number of ways practiced in the art. In one aspect, the candidate bioactive agent is labeled, and binding is determined directly. Where the screening assay is a binding assay, one or more of the molecules may be joined to a label, where the label can directly or indirectly provide a detectable signal. Various labels include radioisotopes, enzymes, fluorescent and chemiluminescent compounds, specific binding molecules, particles, e.g. magnetic particles, and the like. Specific binding molecules include pairs, such as biotin and streptavidin, digoxin and antidigoxin etc. For the specific binding members, the complementary member would normally be labeled with a molecule which allows detection, in

accordance with known procedures. In some embodiments, only one of the components is labeled. Alternatively, more than one component may be labeled with different labels; for example, the HDAC polypeptide may be labeled with one fluorophor and the candidate agent labeled with another

In one embodiment, the candidate bioactive agent is labeled. Labeled candidate bioactive agents are incubated with an HDAC polypeptide for a time sufficient to allow binding, if present. Incubations may be performed at any temperature which facilitates optimal activity, typically between 4ºC and 40ºC. Incubation periods are selected for optimum activity, but may also be optimized to facilitate rapid high throughput screening. Typically between 0.1 and 1 hour is sufficient. Excess reagent is generally removed or washed away. The presence or absence of the labeled component is detected to determine and indicate binding.

A variety of other reagents may be included in the screening assay. Such reagents include, but are not limited to, salts, neutral proteins, e.g. albumin, detergents, etc., which may be used to facilitate optimal protein-protein binding and/or to reduce non-specific or background interactions. In addition, reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc. may be used. Further, the mixture of components in the method may be added in any order that provides for the requisite binding.

Kits are included as an embodiment of the present invention which comprise containers with reagents necessary to screen test compounds. Depending on the design of the test and the types of compounds to be screened, such kits include human HDAC polynucleotide, polypeptide, or peptide and instructions for performing the assay.

Inhibitors of the enzymatic activity of each of the novel HDAC polypeptides can be identified using assays which measure the ability of an agent to inhibit catalytic conversion of a substrate by the HDAC proteins provided by the present invention. For example, the ability of the novel HDAC proteins to deacetylate a histone substrate, such as histone H4, in the

presence and absence of a candidate inhibitor, can be determined using standard enzymatic assays.

A number of methods have been employed in the art for assaying histone deacetylase activity, and can be incorporated in the drug screening assays of the present invention. Preferably, the assay method will employ a labeled acetyl group linked to appropriate histone lysine residues as substrates. In other embodiments, a histone substrate peptide can be labeled with a group whose signal is dependent on the simultaneous presence or absence of an acetyl group, e.g., the label can be a fluorogenic group whose fluorescence is modulated (either quenched or potentiated) by the presence of the acetyl moiety.

Using standard enzymatic analysis, the ability of a test agent (i.e., test compound) to cause a statistically significant change in substrate conversion by a histone deacetylase can be measured, and as desirable, inhibition constants, e.g., K_i values, can be calculated. The histone substrate can be provided as a purified or semi-purified polypeptide or as part of a cell lysate. Likewise, the histone deacetylase can be provided to a reaction mixture as a purified or semi-purified polypeptide, or as a cell lysate. Accordingly, the reaction mixtures can range from reconstituted protein mixtures derived with purified preparations of histones and deacetylases, to mixtures of cell lysates, e.g., by admixing baculovirus lysates containing recombinant histones and deacetylases.

As an example, the histone substrate for assays described herein can be provided by isolation of radiolabeled histones from metabolically labeled cells. Cells such as HeLa cells can be labeled in culture by the addition of [^{3}H]acetate (New England Nuclear) to the culture media. (Hay et al., 1983, *J. Biol. Chem.*, 258:3726-3734). The addition of an HDAC inhibitor, such as butyrate, trapoxin and the like, can be used to increase the abundance of acetylated histones in the cells. Radiolabeled histones can be isolated from the cells by extraction with H_2SO_4 (Marushige et al., 1966, *J. Mol. Biol.*, 15:160-174). Briefly, cells are homogenized in buffer, centrifuged to isolate a nuclear pellet, and the subsequently homogenized nuclear pellet is

centrifuged through sucrose. The resulting chromatin pellet extracted by addition of H_2SO_4 to yield [^{3}H]acetyl-labeled histones. Alternatively, nucleosome preparations containing [^{3}H]acetyl-labeled histones can be isolated from metabolically labeled cells. As known in the art, nucleosomes can be isolated from cell preparations by sucrose gradient centrifugation (e.g., Hay et al., 1983, *J. Biol. Chem.*, 258:3726-3734 and Noll, 1967, *Nature*, 215:360-363), and polynucleosomes can be prepared by NaCl precipitation from micrococcal nuclease digested cells (Hay et al., *supra*).

Similar procedures for isolating labeled histones from other cells types, including yeast, have been described. (See for example, Alonso et al., 1986, *Biochem Biophys Acta*, 866:161-169 and Kreiger et al, 1974, *J. Biol. Chem.*, 249:332 334). Also, histones are generated by recombinant gene expression, and include an exogenous tag (e.g., an HA epitope, a poly(his) sequence, and the like) which facilitates purification from cell extracts. Further, whole nuclei can be isolated from metabolically labeled cells by micrococcal nuclease digestion (Hay et al., *supra*).

The deacetylase substrate can also be provided as an acetylated peptide including a sequence corresponding to the sequence around the specific lysyl residues acetylated on histones, e.g., peptidyl portions of the core histones H2A, H2B, H3, or H4. Such fragments can be produced by cleavage of acetylated histones derived from metabolically labeled cells, e.g., by treatment with proteolytic enzymes or cyanogen bromide (Kreiger et al., *supra*). The acetylated peptide can also be provided by standard solid phase synthesis using acetylated lysine residues (*Id.*).

The activity of a histone deacetylase in assay detection methods involving use of [^{3}H]acetyl-labeled histones is detected by measuring the release of [^{3}H]acetate by standard scintillation techniques. As an illustrative example, a reaction mixture is provided which contains a recombinant HDAC protein suspended in buffer, along with a sample of [^{3}H]acetyl-labeled histones and (optionally) a test compound. The reaction mixture is maintained at a desired temperature and pK such as 22°C at pH 7.8, for several hours, and the reaction is terminated by boiling, or another form of

denaturation. Released [^{3}H]acetate is extracted and counted. For example, the quenched reaction mixture can be acidified with concentrated HCl and used to create a biphasic mixture with ethyl acetate. The resulting two-phase system is thoroughly mixed, centrifuged, and the ethyl acetate phase collected and counted by standard scintillation methods. Other methods for detecting acetate release will be easily recognized by those having skill in the art.

In yet another aspect, the drug screening assay is designed to include a reagent cell recombinantly expressing one or more of a target protein or HDAC protein. The ability of a test agent to alter the activity of the HDAC protein can be detected by analysis of the recombinant cell. For instance, agonists and antagonists of the HDAC biological activity can by detected by scoring for alterations in growth or differentiation (phenotype) of the cell. General techniques for detecting these characteristics are well known, and will vary with respect to the source of the particular reagent cell utilized in any given assay. For example, quantification of cell proliferation in the presence and absence of a candidate agent can be measured by using a number of techniques well known in the art, including simple measurement of population growth curves.

Where an assay involves proliferation in a liquid medium, turbidimetric techniques (i.e. absorbance/transmittance of light of a given wavelength through the sample) can be utilized. For example, in a case in which the reagent cell is a yeast cell, measurement of absorbance of light at a wavelength at between 540 and 600 nm can provide a conveniently fast measure of cell growth. Moreover, the ability of yeast cells to form colonies in solid medium (e.g. agar) can be used to readily score for proliferation. In other embodiments, an HDAC substrate protein, such as a histone, can be provided as a fusion protein which permits the substrate to be isolated from cell lysates and the degree of acetylation detected. Each of these techniques is suitable for high throughput analysis necessary for rapid screening of large numbers of candidate HDAC modulatory agents.

In addition, in assays in which the ability of an agent to cause or reverse a transformed phenotype is being determined, cell growth in solid or semi-solid medium, such as agar, can further aid in establishing whether a mammalian cell is transformed. Visual inspection of the morphology of the reagent cell can also be used to determine whether the biological activity of the targeted HDAC protein has been affected by the added agent. By illustration, the ability of an agent to influence an apoptotic phenotype which is mediated in some way by a recombinant HDAC protein can be assessed by visual microscopy. Similarly, the formation of certain cellular structures as part of normal cell differentiation, such as the formation of neuritic processes, can be visualized under a light microscope.

The nature of the effect of a test agent on a reagent cell can be assessed by measuring levels of expression of specific genes, e.g., by reverse transcription PCR. Another method of scoring for an effect on HDAC activity is by detecting cell-type specific marker expression through immunofluorescent staining. Many such markers are known in the art for which antibodies are readily available. For example, the presence of chondroitin sulfate proteoglycans, as well as type-II collagen, is correlated with cartilage production in chondrocytes, and each can be detected by immunostaining. Similarly, the human kidney differentiation antigen gp160, human aminopeptidase A, is a marker of kidney induction, and the cytoskeletal protein troponin I is a marker of heart induction.

Also, the alteration of expression of a reporter gene construct provided in the reagent cell provides a means of detecting an effect on HDAC activity. For example, reporter gene constructs designed using transcriptional regulatory sequences, e.g. the promoters, for developmentally regulated genes can be used to drive the expression of a detectable marker, such as a luciferase gene. For example, the construct can be prepared using the promoter sequence from a gene expressed in a particular differentiation phenotype.

Pharmaceutical Compositions

A further embodiment of the present invention embraces the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, diluent, or excipient, for any of the above-described therapeutic uses and effects. Such pharmaceutical compositions may comprise HDAC nucleic acid, polypeptide, or peptides, antibodies to HDAC polypeptides or peptides, or fragments thereof, mimetics, agonists (e.g., activators), antagonists (e.g., inhibitors, blockers) of the HDAC polypeptide, peptide, or polynucleotide. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical (or physiologically compatible) carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, hormones, or biological response modifiers. Preferred are compositions comprising one or more HDAC inhibitors.

The pharmaceutical compositions for use in the present invention can be administered by any number of routes including, but not limited to, parenteral oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, ophthalmic, enteral, topical, sublingual, vaginal, or rectal means.

Transdermal patches have the added advantage of providing controlled delivery of a compound of the present invention to the body. Such dosage forms can be made by dissolving or dispersing a deacetylase inhibitor in the proper medium. Absorption enhancers can also be used to increase the flux of the deacetylase inhibitor across the skin. The rate of such flux can be controlled by either providing a rate controlling membrane or dispersing the deacetylase inhibitor in a polymer matrix or gel.

Ophthalmic formulations, eye ointments, powders, solutions and the like, are also contemplated as being within the scope of this invention.

In addition to the active ingredients (i.e., an HDAC antagonist compound), the pharmaceutical compositions may contain suitable

pharmaceutically acceptable carriers or excipients comprising auxiliaries which facilitate processing of the active compounds into preparations that can be used pharmaceutically. Further details on techniques for formulation and administration are provided in the latest edition of *Remington's Pharmaceutical Sciences* (Maack Publishing Co., Easton, Pa.).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained by the combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethylcellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth, and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a physiologically acceptable salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with physiologically suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification, or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, scaled capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain

active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. In addition, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyloleate or triglycerides, or liposomes. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants or permeation agents that are appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants and permeation enhancers are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, and the like. Salts tend to be more soluble in aqueous solvents, or other protonic solvents, than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to

5.5, combined with a buffer prior to use. After the pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of an HDAC inhibitor compound, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose or amount is well within the capability of those skilled in the art. For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., using neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used and extrapolated to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example, an HDAC inhibitor or antagonist compound, antibodies to an HDAC polypeptide or peptide, agonists of HDAC polypeptides, which ameliorates, reduces, or eliminates the symptoms or the condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED_{50} (the dose therapeutically effective in 50% of the population) and LD_{50} (the dose lethal to 50% of the population). The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the ratio, LD_{50}/ED_{50}. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used in determining a range of dosages for human use. Preferred dosage contained in a pharmaceutical composition is within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, who will consider the factors related to the individual requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the individual's disease state, general health of the patient, age, weight, and gender of the patient, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. As a general guide, long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks, depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms (μg), up to a total dose of about 1 gram (g), depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and is generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, and the like.

<u>Assays and Diagnostics</u>

In another embodiment of the present invention, antibodies which specifically bind to the HDAC polypeptides or peptides of the present invention may be used for the diagnosis of conditions or diseases characterized by expression (or overexpression) of an HDAC polynucleotide or polypeptide, or in assays to monitor patients being treated modulatory compounds of HDAC polypeptides, or, for example, HDAC antagonists or inhibitors. The antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for use in therapeutic methods. Diagnostic assays for the HDAC polypeptides include methods which utilize the antibody and a label to detect the protein in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a

reporter molecule. A wide variety of reporter molecules which are known in the art may be used, several of which are described above.

Several assay protocols including ELISA, RIA, and FACS for measuring an HDAC polypeptide or peptide are known in the art and provide a basis for diagnosing altered or abnormal levels of HDAC polypeptide expression. Normal or standard values for HDAC polypeptide expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to HDAC polypeptide or peptide under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods; photometric means are preferred. Quantities of HDAC polypeptide or peptide expressed in subject sample, control sample, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In one embodiment of the present invention, anti-HDAC antibodies (e.g., anti-HDAC9c antibodies) can be used in accordance with established methods to detect the presence of specific cancers or tumors, such as breast or prostate cancers or tumors. Representative cancers and cancer types are listed above.

According to another embodiment of the present invention, the polynucleotides encoding the novel HDAC polypeptides may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify HDAC-encoding nucleic acid expression in biopsied tissues in which expression (or under- or overexpression) of HDAC polynucleotide may be correlated with disease. The diagnostic assay may be used to distinguish between the absence, presence, and excess expression of HDAC, and to monitor regulation of HDAC polynucleotide levels during therapeutic treatment or intervention.

In a related aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding an HDAC polypeptide, or closely related molecules, may be used to identify nucleic acid sequences which encode an HDAC polypeptide. The specificity of the probe, whether it is made from a highly specific region, e.g., about 8 to 10 or 12 or 15 contiguous nucleotides in the 5' regulatory region, or a less specific region, e.g., especially in the 3' coding region, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences encoding the HDAC polypeptide, alleles thereof, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50%, preferably at least 80%, of the nucleotides encoding an HDAC polypeptide. The hybridization probes of this invention may be DNA or RNA and may be derived from the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, or SEQ ID NO:96, or from genomic sequence including promoter, enhancer elements, and introns of the naturally occurring HDAC protein.

The nucleotide sequences of the novel HDAC genes presented herein will further allow for the generation of probes and primers designed for use in identifying and/or cloning HDAC homologs in other cell types, e.g. from other tissues, as well as HDAC homologs from other organisms. For example, the present invention also provides a probe/primer comprising a substantially purified oligonucleotide, which oligonucleotide comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or anti-sense sequence selected from the group consisting of HDAC SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, or SEQ ID NO:96, or naturally occurring mutants thereof. Primers based on the nucleic acid represented in SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, or SEQ ID NO:96, or as presented in the tables herein, can be used in PCR

reactions to clone HDAC homologs. Likewise, probes based on the HDAC sequences provided herein can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. The probe preferably comprises a label moiety attached thereto and is able to be detected, e.g., the label moiety is selected from radioisotopes, fluorescent compounds, chemiluminescent compounds, enzymes, enzyme co-factors, and the like.

Such probes can also be used as a part of a diagnostic test kit for identifying cells or tissue which mis-express an HDAC protein, such as by measuring a level of an HDAC encoding nucleic acid in a sample of cells from a patient; e.g., detecting HDAC mRNA levels, or determining whether a genomic HDAC gene has been mutated or deleted. To this end, nucleotide probes can be generated from the HDAC sequences herein which facilitate histological screening of intact tissue and tissue samples for the presence (or absence) of HDAC-encoding transcripts. Similar to the diagnostic uses of anti-HDAC antibodies, the use of probes directed to HDAC messages, or to genomic HDAC sequences, can be used for both predictive and therapeutic evaluation of allelic mutations which might be manifest in, for example, neoplastic or hyperplastic disorders (e.g. unwanted cell growth), or the abnormal differentiation of tissue. Used in conjunction with immunoassays as described herein, the oligonucleotide probes can help facilitate the determination of the molecular basis for a developmental disorder which may involve some abnormality associated with expression (or lack thereof) of an HDAC protein. For instance, variation in polypeptide synthesis can be differentiated from a mutation in a coding sequence.

Accordingly, the present invention provides a method for determining if a subject is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. Such a method can be generally characterized as comprising detecting, in a sample of cells from a subject, the presence or absence of a genetic lesion characterized by at least one of (i) an alteration affecting the integrity of a gene or nucleic acid sequence encoding an HDAC polypeptide, or (ii) the mis-expression of an HDAC gene. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one

of (i) a deletion of one or more nucleotides from an HDAC gene, (ii) an addition of one or more nucleotides to an HDAC gene, (iii) a substitution of one or more nucleotides of an HDAC gene, (iv) a gross chromosomal rearrangement of an HDAC gene, (v) a gross alteration in the level of a messenger RNA transcript of an HDAC gene, (vii) aberrant modification of an HDAC gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild type splicing pattern of a messenger RNA transcript of an HDAC gene, (viii) a non-wild type level of an HDAC polypeptide, and (ix) inappropriate post-translational modification of an HDAC polypeptide. Accordingly, the present invention provides a large number of assay techniques for detecting lesions in an HDAC gene, and importantly, provides the ability to distinguish between different molecular causes underlying HDAC-dependent aberrant cell growth, proliferation and/or differentiation.

Methods for producing specific hybridization probes for DNA encoding the HDAC polypeptides include the cloning of nucleic acid sequence that encodes the HDAC polypeptides, or HDAC derivatives, into vectors for the production of mRNA probes. Such vectors are known in the art, commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of detector/reporter groups, e.g., radionuclides such as ^{32}P or ^{35}S, or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/ biotin coupling systems, and the like.

The polynucleotide sequences encoding the HDAC polypeptides may be used in Southern or Northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dip stick, pin, ELISA or chip assays utilizing fluids or tissues from patient biopsies to detect the status of, e.g., levels or overexpression of HDAC, or to detect altered HDAC expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding the HDAC polypeptides may be useful in assays that detect activation or induction of various tumors, neoplasms or cancers. The nucleotide sequences encoding

the HDAC polypeptides may be labeled by standard methods, and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the biopsied or extracted sample is significantly altered from that of a comparable control sample, the nucleotide sequence has hybridized with nucleotide sequence present in the sample, and the presence of altered levels of nucleotide sequence encoding the HDAC polypeptides in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or in monitoring the treatment of an individual patient.

In one embodiment of the present invention, HDAC (e.g., HDAC9c) nucleic acids can be used in accordance with established methods to detect the presence of specific cancers or tumors, such as breast or prostate cancers or tumors. Representative cancers and cancer types are listed herein above.

To provide a basis for the diagnosis of disease associated with HDAC expression, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, which encodes an HDAC polypeptide, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for disease. Deviation between standard and subject (patient) values is used to establish the presence of disease.

Once disease is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to evaluate whether the level of expression in the patient begins to approximate that which is

observed in a normal individual. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the nucleic acid sequences encoding the novel HDAC polypeptides may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically, or produced from a recombinant source. Oligomers will preferably comprise two nucleotide sequences, one with sense orientation (5'→3') and another with antisense (3'→5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantification of closely related DNA or RNA sequences.

Methods suitable for quantifying the expression of HDAC include radiolabeling or biotinylating nucleotides, co-amplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (P.C. Melby et al., 1993, *J. Immunol. Methods*, 159:235-244; and C. Duplaa et al., 1993, *Anal. Biochem.*, 229-236). The speed of quantifying multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantification.

In another embodiment of the present invention, oligonucleotides, or longer fragments derived from the HDAC polynucleotide sequences described herein, may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously (to

produce a transcript image), and to identify genetic variants, mutations and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disease, to diagnose disease, and to develop and monitor the activities of therapeutic agents. In a particular aspect, the microarray is prepared and used according to the methods described in WO 95/11995 (Chee et al.); D.J. Lockhart et al., 1996, *Nature Biotechnology*, 14:1675-1680; and M. Schena et al., 1996, *Proc. Natl. Acad. Sci. USA*, 93:10614-10619). Microarrays are further described in U.S. Patent No. 6,015,702 to P. Lal et al.

In another embodiment of this invention, a nucleic acid sequence which encodes one or more of the novel HDAC polypeptides may also be used to generate hybridization probes which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial PI constructions, or single chromosome cDNA libraries, as reviewed by C.M. Price, 1993, *Blood Rev.*, 7:127-134 and by B.J. Trask, 1991, *Trends Genet.*, 7:149-154.

In another embodiment of the present invention, an HDAC polypeptide, its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between an HDAC polypeptide, or portion thereof, and the agent being tested, may be measured utilizing techniques commonly practiced in the art and as described above.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the protein of interest as described in WO 84/03564. In this method, as applied to HDAC protein, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with an HDAC polypeptide, or fragments

thereof, and washed. Bound HDAC polypeptide is then detected by methods well known in the art. Purified HDAC polypeptide can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

Other screening and small molecule (e.g., drug) detection assays which involve the detection or identification of small molecules that can bind to a given protein, i.e., an HDAC protein, are encompassed by the present invention. Particularly preferred are assays suitable for high throughput screening methodologies. In such binding-based screening or detection assays, a functional assay is not typically required. All that is needed is a target protein, preferably substantially purified, and a library or panel of compounds (e.g., ligands, drugs, small molecules) to be screened or assayed for binding to the protein target. Preferably, most small molecules that bind to the target protein will modulate activity in some manner, due to preferential, higher affinity binding to functional areas or sites on the protein.

An example of such an assay is the fluorescence based thermal shift assay (3-Dimensional Pharmaceuticals, Inc., 3DP, Exton, PA) as described in U.S. Patent Nos. 6,020,141 and 6,036,920 to Pantoliano et al.; see also, J. Zimmerman, 2000, *Gen. Eng. News* 20(8)). The assay allows the detection of small molecules (e.g., drugs, ligands) that bind to expressed, and preferably purified, HDAC polypeptide based on affinity of binding determinations by analyzing thermal unfolding curves of protein-drug or ligand complexes. The drugs or binding molecules determined by this technique can be further assayed, if desired, by methods, such as those described herein, to determine if the molecules affect or modulate function or activity of the target protein.

In a further embodiment of this invention, competitive drug screening assays can be used in which neutralizing antibodies capable of binding an HDAC polypeptide specifically compete with a test compound for binding to HDAC polypeptide. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with an HDAC polypeptide.

In yet another of its aspects, the present invention provides the identification of compounds with optimum therapeutic indices, or drugs or compounds which have therapeutic indices more favorable than known HDAC inhibitors, such as trapoxin, tichostatin, sodium butyrate, and the like. The identification of such compounds can be made by the use of differential screening assays which detect and compare drug mediated inhibition of deacetylase activity between two or more different HDAC-like enzymes, or which compare drug mediated inhibition of formation of complexes involving two or more different types of HDAC-like proteins.

For example, an assay can be designed for side-by side comparison of the effect of a test compound on the deacetylase activity or protein interactions of tissue-type specific HDAC proteins. Given the apparent diversity of HDAC proteins, it is probable that different functional HDAC activities, or HDAC complexes, exist and in certain instances, are localized to particular tissue or cell types. Thus, test compounds can be screened to identify agents that are able to inhibit the tissue-specific formation of only a subset of the possible repertoire of HDAC/regulatory protein complexes, or which preferentially inhibit certain HDAC enzymes. For instance, an "interaction trap assay" can be derived using two or more different human HDAC "bait" proteins, while the "fish" protein is constant in each, e.g., a human RbAp48 construct. Running the interaction trap side- by-side permits the detection of agents which have a greater effect (e.g., statistically significant) on the formation of one of the HDAC/RbAp48 complexes than on the formation of the other HDAC complexes. (See, e.g., WO 97/35990).

Similarly, differential screening assays can be used to exploit the difference in protein interactions and/or catalytic mechanisms of mammalian HDAC proteins and yeast RPD3 proteins, for example, in order to identify agents which display a statistically significant increase in specificity for inhibiting the yeast enzyme relative to the mammalian enzyme. Thus, lead compounds which act specifically on pathogens, such as fungus involved in mycotic infections, can be developed. By way of illustration, assays can be used to screen for agents which may ultimately be useful for inhibiting at least

one fungus implicated in pathologies such as candidiasis, aspergillosis, mucomycosis, blastomycosis, geotrichosis, cryptococcosis, chromoblastomycosis, coccidiomycosis, conidiosporosis, histoplasmosis, maduromycosis, rhinosporidosis, nocaidiosis, para actinomycosis, penicilliosis, monoliasis, or sporotrichosis.

As an example, if the mycotic infection to which treatment is desired is candidiasis, the described assay can involve comparing the relative effectiveness of a test compound on inhibiting the deacetylase activity of a mammalian HDAC protein with its effectiveness in inhibiting the deacetylase activity of an RPD3 homolog that has been cloned from yeast selected from the group consisting of *Candida albicans, Candida stellatoidea, Candida tropicalis, Candida parapsilosis, Candida krusei, Candida pseudotropicalis, Candida quillermondii*, or *Candida rugosa.* Such an assay can also be used to identify anti-fungal agents which may have therapeutic value in the treatment of aspergillosis by selectively targeting RPD3 homologs cloned from yeast such as *Aspergillus fumigatus, Aspergillus flavus, Aspergillus niger, Aspergillus nidulans*, or *Aspergillus terreus*. Where the mycotic infection is muco-mycosis, the RPD3 deacetylase can be derived from yeast such as *Rhizopus arrhizus, Rhizopus oryzae, Absidja corymbiera, Absidia ramosa*, or *Mucor pusillus.*

Sources of other RPD3 activities for comparison with a mammalian HDAC activity include the pathogen *Pneumocystis carinii.*

In addition to such HDAC therapeutic uses, anti-fungal agents developed from such differential screening assays can be used, for example, as preservatives in foodstuff, feed supplement for promoting weight gain in livestock, or in disinfectant formulations for treatment of non-living matter, e.g., for decontaminating hospital equipment and rooms. In a similar fashion, side by side comparison of the inhibition of a mammalian HDAC protein and an insect HDAC-related protein, will permit selection of HDAC inhibitors which are capable of discriminating between the human/mammalian and insect enzymes. Accordingly, the present invention envisions the use and

formulations of HDAC therapeutics in insecticides, such as for use in management of insects like the fruit fly.

In yet another embodiment, certain of the subject HDAC inhibitors can be selected on the basis of inhibitory specificity for plant HDAC-related activities relative to the mammalian enzyme. For example, a plant HDAC-related protein can be disposed in a differential screen with one or more of the human enzymes to select those compounds of greatest selectivity for inhibiting the plant enzyme. Thus, the present invention specifically contemplates formulations of HDAC inhibitors for agricultural applications, such as in the form of a defoliant or the like.

In many drug screening programs that test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be rapidly generated to permit the quick development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. In addition, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in an *in vitro* system, since the assay is focused primarily on the effect of the drug on the molecular target which may be manifest in an alteration of binding affinity with upstream or downstream elements.

Accordingly, in an exemplary screening assay, a reaction mixture is generated to include an HDAC polypeptide, compound(s) of interest, and a "target polypeptide", e.g., a protein, which interacts with the HDAC polypeptide, whether as a substrate or by some other protein-protein interaction. Exemplary target polypeptides include histones, RbAp48 polypeptides, p53 polypeptides, and/or combinations thereof, or with other transcriptional regulatory proteins (such as myc, max, etc.). Detection and quantification of complexes containing the HDAC protein provide a means for determining a compound's efficacy at inhibiting (or potentiating) complex formation between the HDAC and the target polypeptide. The efficacy of the

compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, isolated and purified HDAC polypeptide is added to a composition containing the target polypeptide and the formation of a complex is quantified in the absence of the test compound.

Complex formation between an HDAC polypeptide and the target polypeptide may be detected by a variety of techniques. Modulation of the formation of complexes can be quantified using, for example, detectably labeled proteins such as radiolabeled, fluorescently labeled, or enzymatically labeled HDAC polypeptides, by immunoassay, by chromatography, or by detecting the intrinsic activity of the acetylase.

Transgenics and Knock Outs

The present invention further encompasses transgenic non-human mammals, preferably mice, that comprise a recombinant expression vector harboring a nucleic acid sequence that encodes a human HDAC (e.g., SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, or SEQ ID NO:95).

Transgenic non-human mammals useful to produce recombinant proteins are well known to the skilled practitioner, as are the expression vectors necessary and the techniques for generating transgenic animals. Generally, the transgenic animal comprises a recombinant expression vector in which the nucleotide sequence that encodes a human HDAC is operably linked to a tissue specific promoter whereby the coding sequence is only expressed in that specific tissue. For example, the tissue specific promoter can be a mammary cell specific promoter and the recombinant protein so expressed is recovered from the animal's milk.

The transgenic animals, particularly transgenic mice, containing a nucleic acid molecule which encodes a novel human HDAC may be used as animal models for studying *in vivo* the overexpression of HDAC and for use in drug evaluation and discovery efforts to find compounds effective to inhibit or modulate the activity of HDAC, such as for example compounds for treating

disorders, diseases, or conditions related to cell proliferation and neoplastic cell growth, for example. One having ordinary skill in the art using standard techniques, such as those taught in U.S. Patent No. 4,873,191, issued Oct. 10, 1989 to Wagner and in U.S. Patent No. 4,736,866, issued April 12, 1988 to Leder, can produce transgenic animals which produce human HDAC, and use the animals in drug evaluation and discovery projects.

The transgenic non-human animals according to this aspect of the present invention can express a heterologous HDAC-encoding gene, or which have had one or more genomic HDAC genes disrupted in at least one of the tissue or cell types of the animal. Accordingly, the invention features an animal model for developmental diseases, which animal has one or more HDAC alleles which are improperly expressed. For example, a mouse can be bred which has one or more HDAC alleles deleted or otherwise rendered inactive. Such a mouse model can then be used to study disorders arising from improperly expressed HDAC genes, as well as for evaluating potential therapies for similar disorders.

Another aspect of transgenic animals are those animals which contain cells harboring an HDAC transgene according to the present invention and which preferably express an exogenous HDAC protein in one or more cells in the animal. An HDAC transgene can encode the wild-type form of the protein, or can encode homologs thereof, including both agonists and antagonists, as well as antisense constructs. Preferably, the expression of the transgene is restricted to specific subsets of cells, tissues or developmental stages utilizing, for example, cis-acting sequences that control expression in the desired pattern. According to the invention, such mosaic expression of an HDAC protein can be essential for many forms of lineage analysis and can also provide a means to assess the effects of, for example, lack of HDAC expression which might grossly alter development in small portions of tissue within an otherwise normal embryo. Toward this end, tissue specific regulatory sequences and conditional regulatory sequences can be used to control the expression of the transgene in certain spatial patterns. Moreover,

temporal patterns of expression can be provided by, for example, conditional recombination systems or prokaryotic transcriptional regulatory sequences.

Genetic techniques which allow for the expression of transgenes can be regulated via site-specific genetic manipulation *in vivo* are known to those skilled in the art. For instance, genetic systems are available which permit the regulated expression of a recombinase that catalyzes the genetic recombination of a target sequence. The phrase "target sequence" in this instance refers to a nucleotide sequence that is genetically recombined by a recombinase. The target sequence is flanked by recombinase recognition sequences and is generally either excised or inverted in cells expressing recombinase activity. Recombinase catalyzed recombination events can be designed such that recombination of the target sequence results in either the activation or repression of expression of one of the present HDAC proteins.

For example, excision of a target sequence which interferes with the expression of a recombinant HDAC gene, such as one which encodes an antagonistic homolog or an antisense transcript, can be designed to activate the expression of that gene. This interference with expression of an encoded product can result from a variety of mechanisms, such as spatial separation of the HDAC gene from the promoter element, or an internal stop codon. Moreover, the transgene can be made so that the coding sequence of the gene is flanked by recombinase recognition sequences and is initially transfected into cells in a 3' to 5' orientation with respect to the promoter element. In this case, inversion of the target sequence will reorient the subject gene by placing the 5' end of the coding sequence in an orientation with respect to the promoter element which allows for promoter driven transcriptional activation.

Illustratively, transgenic non-human animals are produced by introducing transgenes into the germline of the non-human animal. Embryonic target cells at various developmental stages can be used to introduce transgenes. Different methods are used depending on the stage of development of the embryonic target cell. The zygote is a preferred target for micro-injection.

In the mouse, the male pronucleus reaches the size of approximately 20 micrometers in diameter which allows reproducible injection of 1-2pl of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host gene before the first cleavage (e.g., Brinster et al., 1985, *Proc. Natl. Acad. Sci. USA*, 82:4438-4442). As a consequence, all cells of the transgenic non-human animal will carry the incorporated transgene. This will generally also be reflected in the efficient transmission of the transgene to offspring of the founder mice since 50% of the germ cells will harbor the transgene. Microinjection of zygotes is the preferred method for incorporating HDAC transgenes.

In addition, retroviral infection can also be used to introduce HDAC transgenes into a non human animal. The developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres are targets for retroviral infection (R. Jaenisch, 1976, *Proc. Natl. Acad. Sci. USA.,* 73:1260-1264). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (Manipulating the Mouse Embryo, Hogan eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1986). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahner et al., 1985, *Proc. Natl. Acad. Sci. USA.*, 82:6927 6931; Van der Putten et al., 1985, *Proc. Natl. Acad. Sci. USA.*, 82:6148-6152). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Stewart et al., 1987, *EMBO J.*, 6:383-388).

Alternatively, infection can be performed at a later developmental stage. For example, virus or virus-producing cells can be injected into the blastocoele (e.g., Jahner et al., 1982, *Nature*, 298:623-628). Most of the founder animals win be mosaic for the transgene, because incorporation occurs only in the subset of cells which formed the transgenic non-human animal. Further, the founders may contain various retroviral insertions of the transgene at different positions in the genome which generally will segregate in the offspring. It is also possible to introduce transgenes into the germline

by intrauterine retroviral infection of the midgestation embryo (Jahner et al., 1982, *supra*).

A third type of target cell for transgene introduction is the embryonic stem cell (ES). ES cells are obtained from pre-implantation embryos that are cultured *in vitro* and fused with embryos (Evans et al., 1981, *Nature*, 292:154-156; Bradley et al., 1984, *Nature*, 309:255-258; Gossler et al., 1986, *Proc. Natl. Acad. Sci. USA.*, 83:9065-9069; and Robertson et al., 1986, *Nature*, 322:445-448). Cultured ES cell lines are available. Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retrovirus-mediated transduction. Transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The ES cells then colonize the embryo and contribute to the germ line of the resulting chimeric animal. See, e.g., R. Jaenisch, 1988, *Science*, 240:1468-1474.

Methods for making HDAC knock-out animals, or disruption transgenic animals are also generally known. See, for example, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Recombinase dependent knockouts can also be generated, e.g. by homologous recombination, to insert recombinase target sequences flanking portions of an endogenous HDAC gene, such that tissue specific and/or temporal control of inactivation of an HDAC gene sequence or allele can be controlled as above.

In knock-outs, transgenic mice may be generated which are homozygous for a mutated, non-functional HDAC gene which is introduced into the animals using well known techniques. Surviving knock-out mice produce no functional HDAC and thus are useful to study the function of HDAC. Furthermore, the mice may be used in assays to study the effects of test compounds in HDAC deficient animals. For instance, HDAC-deficient mice can be used to determine if, how and to what extent HDAC inhibitors will effect the animal and thus address concerns associated with inhibiting the activity of the molecule.

More specifically, methods of generating genetically deficient knock-out mice are well known and are disclosed in M.R. Capecchi, 1989, *Science*,

244:1288-1292 and P. Li et al., 1995, *Cell*, 80:401-411. For example, a human HDAC cDNA clone can be used to isolate a murine HDAC genomic clone. The genomic clone can be used to prepare an HDAC targeting construct which can disrupt the HDAC gene in the mouse by homologous recombination. The targeting construct contains a non-functioning portion of an HDAC gene which inserts in place of the functioning portion of the native mouse gene. The non-functioning insert generally contains an insertion in the exon that encodes the active region of the HDAC polypeptide. The targeting construct can contain markers for both positive and negative selection. The positive selection marker allows for the selective elimination of cells which do not carry the marker, while the negative selection marker allows for the elimination of cells that carry the marker.

For example, a first selectable marker is a positive marker that will allow for the survival of cells carrying it. In some instances, the first selectable marker is an antibiotic resistance gene, such as the neomycin resistance gene, which can be placed within the coding sequence of a novel HDAC gene to render it non-functional, while at the same time rendering the construct selectable. The antibiotic resistance gene is within the homologous region which can recombine with native sequences. Thus, upon homologous recombination, the non-functional and antibiotic resistance selectable gene sequences will be taken up. Knock-out mice may be used as models for studying inflammation-related disorders and screening compounds for treating these disorders.

The targeting construct also contains a second selectable marker which is a negative selectable marker. Cells with the negative selectable marker will be eliminated. The second selectable marker is outside the recombination region. Thus, if the entire construct is present in the cell, both markers will be present. If the construct has recombined with native sequences, the first selectable marker will be incorporated into the genome and the second will be lost. The herpes simplex virus thymidine kinase (HSV tk) gene is an example of a negative selectable marker which can be used as

a second marker to eliminate cells that carry it. Cells with the HSV tk gene are selectively killed in the presence of gangcyclovir.

Cells are transfected with targeting constructs and then selected for the presence of the first selection marker and the absence of the second. Constructs / DNA are then injected into the blastocyst stage and implanted into pseudopregnant females. Chimeric offspring which are capable of transferring the recombinant genes in their germline are selected, mated and their offspring examined for heterozygous carriers of the recombined genes. Mating of the heterozygous offspring can then be used to generate fully homozygous offspring which constitute HDAC-deficient knock-out mice.

Embodiments of the Invention

- An isolated polynucleotide encoding a histone deacetylase polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, and SEQ ID NO:95.
- An isolated polynucleotide encoding an amino acid sequence selected from the group consisting of:

  a. an amino acid sequence comprising residues 1009-1069 of SEQ ID NO:87; and

  b. an amino acid sequence comprising residues 720-780 of SEQ ID NO:93.
- An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, and SEQ ID NO:96.
- An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

  a. a nucleotide sequence which is at least 60% identical to SEQ ID NO:1;

  b. a nucleotide sequence which is at least 60% identical to SEQ ID NO:12;

  c. a nucleotide sequence which is at least 60% identical to SEQ ID NO:19;

d. a nucleotide sequence which is at least 67.8% identical to SEQ ID NO:88;

e. a nucleotide sequence which is at least 70% identical to SEQ ID NO:94;

f. a nucleotide sequence which is at least 59.8% identical to SEQ ID NO:96;

g. a nucleotide sequence which is at least 94.4% identical to nucleotides 1 to 3207 of SEQ ID NO:88;

h. a nucleotide sequence which is at least 55.4% identical to nucleotides 307 to 1791 of SEQ ID NO:96.

i. a nucleotide sequence comprising nucleotides 1 to 3207 of SEQ ID NO:88;

j. a nucleotide sequence comprising nucleotides 1 to 2340 of SEQ ID NO:94;

k. a nucleotide sequence comprising nucleotides 307 to 1791 of SEQ ID NO:96;

l. a nucleotide sequence comprising nucleotides 4 to 3207 of SEQ ID NO:88 wherein said nucleotides encode amino acids 2 to 1069 of SEQ ID NO:87 lacking the start methionine; and

m. a nucleotide sequence comprising nucleotides 310 to 1791 of SEQ ID NO:96 wherein said nucleotides encode amino acids 2 to 495 of SEQ ID NO:95 lacking the start methionine.

- An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

a. a nucleotide sequence comprising at least 25 contiguous nucleotides of SEQ ID NO1;

b. a nucleotide sequence comprising at least 25 contiguous nucleotides of SEQ ID NO:12;

c. a nucleotide sequence comprising at least 25 contiguous nucleotides of SEQ ID NO:19;

d. a nucleotide sequence comprising at least 2755 contiguous nucleotides of SEQ ID NO:88;

e. a

nucleotide sequence comprising at least 2160 contiguous nucleotides of SEQ ID NO:94; f. a nucleotide sequence comprising at least 1195 contiguous nucleotides of SEQ ID NO:96; g. a nucleotide sequence comprising at least 183 contiguous nucleotides of SEQ ID NO:88; and h. a nucleotide sequence comprising at least 17 contiguous nucleotides of SEQ ID NO:96.

- An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

  a. a nucleotide sequence comprising nucleotides 3024-4467 of SEQ ID NO:88;

  b. a nucleotide sequence comprising nucleotides 2156-3650 of SEQ ID NO:94;

  c. a nucleotide sequence comprising nucleotides 1174-3391 of SEQ ID NO:96;

  d. a nucleotide sequence comprising nucleotides 3024-3207 of SEQ ID NO:88; and

  e. a nucleotide sequence comprising nucleotides 1174-1791 of SEQ ID NO:96.

- An primer comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:24-27, SEQ ID NO:28-35, SEQ ID NO:39-46, SEQ ID NO:47-62, SEQ ID NO:65-66, SEQ ID NO:67-74, SEQ ID NO:75-82, and SEQ ID NO:104-105.
- A probe comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:36, SEQ ID NO:63-64, SEQ ID NO:83-86, SEQ ID NO92, and SEQ ID NO:101-103.
- A cell line comprising the isolated polynucleotide according to any one of the preceding embodiments.
- A gene delivery vector comprising the isolated polynucleotide according to any one of the preceding embodiments.

- An expression vector comprising the isolated polynucleotide according to any one of the preceding embodiments.
- A host cell comprising the expression vector according to any one of the preceding embodiments, wherein the host cell is selected from the group consisting of bacterial, yeast, insect, mammalian, and human cells.
- An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, and SEQ ID NO:95.
- An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

    a. an amino acid sequence which is at least 72% identical to SEQ ID NO:2;

    b. an amino acid sequence which is at least 79% identical to SEQ ID NO:4;

    c. an amino acid sequence which is at least 70% identical to SEQ ID NO:5;

    d. an amino acid sequence which is at least 94.2% identical to SEQ ID NO:87;

    e. an amino acid sequence which is at least 95% identical to SEQ ID NO:93; and

    f. an amino acid sequence which is at least 55.3% identical to SEQ ID NO:95.
- An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

    a. an amino acid sequence comprising at least 8 contiguous amino acids of SEQ ID NO:2;

    b. an amino acid sequence comprising at least 8 contiguous amino acids of SEQ ID NO:4;

    c. an amino acid sequence comprising at least 8 contiguous amino acids of SEQ ID NO:5;

    d. an amino acid sequence comprising at least 920 contiguous amino acids of SEQ ID NO:87;

    e. an amino acid

sequence comprising at least 720 contiguous amino acids of SEQ ID NO:93; and f. an amino acid sequence comprising at least 400 contiguous amino acids of SEQ ID NO:95.

- An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

  a. an amino acid sequence comprising residues 1009-1069 of SEQ ID NO:87; and

  b. an amino acid sequence comprising residues 720-780 of SEQ ID NO:93.

- An isolated fusion protein comprising the isolated polypeptide according to any one of the preceding embodiments.
- An antibody which binds specifically to the isolated polypeptide according to any one of the preceding embodiments, wherein the antibody is selected from the group consisting of polyclonal and monoclonal antibodies.
- An antibody which binds specifically to the isolated fusion protein according to any one of the preceding embodiments.
- An antisense polynucleotide comprising a nucleotide sequence that is complementary to at least 20 contiguous nucleotides of the isolated polynucleotide according to any one of the preceding embodiments.
- An antisense polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:36, SEQ ID NO:63-64, and SEQ ID NO:83-86.
- An expression vector comprising the antisense polynucleotide according to any one of the preceding embodiments.
- A pharmaceutical composition comprising the monoclonal antibody according to any one of the preceding embodiments, and a physiologically acceptable carrier, diluent, or excipient.
- A pharmaceutical composition comprising the antisense polynucleotide according to any one of the preceding embodiments and a physiologically acceptable carrier, diluent, or excipient.

- A pharmaceutical composition comprising the expression vector according to any one of the preceding embodiments, and a physiologically acceptable carrier, diluent, or excipient.
- A pharmaceutical composition comprising the gene delivery vector according to any one of the preceding embodiments, and a physiologically acceptable carrier, diluent, or excipient.
- A pharmaceutical composition comprising the host cell according to any one of the preceding embodiments, and a physiologically acceptable carrier, diluent, or excipient.
- A pharmaceutical composition comprising the modulating agent according to any one of the following embodiments, and a physiologically acceptable carrier, diluent, or excipient.
- A method of treating cancer comprising administering the pharmaceutical composition according to any one of the preceding embodiments in an amount effective for treating the cancer.

  In various aspects, the cancer is selected from the group consisting of bladder cancer, lung cancer, breast cancer, colon cancer, rectal cancer, endometrial cancer, ovarian cancer, head and neck cancer, prostate cancer, and melanoma.

  In other aspects, the breast cancer is selected from the group consisting of ductal carcinoma *in situ*, intraductal carcinoma lobular carcinoma *in situ*, papillary carcinoma, and comedocarcinoma, adenocarcinomas, and carcinomas, such as infiltrating ductal carcinoma, infiltrating lobular carcinoma, infiltrating ductal and lobular carcinoma, medullary carcinoma, mucinous carcinoma, comedocarcinoma, Paget's Disease, papillary carcinoma, tubular carcinoma, and inflammatory carcinoma.

  In further aspects, the prostate cancer is selected from the group consisting of adenocarcinomas and sarcomas, and pre-cancerous conditions, such as prostate intraepithelial neoplasia.
- A method of diagnosing a cancer comprising:

  a. incubating the isolated polynucleotide according to any

one of the preceding embodiments with a biological sample under conditions to allow the isolated polynucleotide to amplify a polynucleotide in the sample to produce a amplification product; and

b. measuring levels of amplification product formed in (a), wherein an alteration in these levels compared to standard levels indicates diagnosis of the cancer.

In various aspects, the cancer is selected from the group consisting of bladder cancer, lung cancer, breast cancer, colon cancer, rectal cancer, endometrial cancer, ovarian cancer, head and neck cancer, prostate cancer, and melanoma. In other aspects, the breast cancer is selected from the group consisting of ductal carcinoma *in situ*, intraductal carcinoma lobular carcinoma *in situ*, papillary carcinoma, and comedocarcinoma, adenocarcinomas, and carcinomas, such as infiltrating ductal carcinoma, infiltrating lobular carcinoma, infiltrating ductal and lobular carcinoma, medullary carcinoma, mucinous carcinoma, comedocarcinoma, Paget's Disease, papillary carcinoma, tubular carcinoma, and inflammatory carcinoma. In further aspects, the prostate cancer is selected from the group consisting of adenocarcinomas and sarcomas, and pre-cancerous conditions, such as prostate intraepithelial neoplasia.

- A method of diagnosing cancer comprising:

a. contacting the antibody according to any one of the preceding embodiments with a biological sample under conditions to allow the antibody to associate with a polypeptide in the sample to form a complex; and

b. measuring levels of complex formed in (a), wherein an alteration in these levels compared to standard levels indicates diagnosis of the cancer.

In various aspects, the cancer is selected from the group consisting of bladder cancer, lung cancer, breast cancer, colon cancer, rectal cancer, endometrial cancer, ovarian cancer, head and neck cancer,

prostate cancer, and melanoma.

In other aspects, the breast cancer is selected from the group consisting of ductal carcinoma *in situ*, intraductal carcinoma lobular carcinoma *in situ*, papillary carcinoma, and comedocarcinoma, adenocarcinomas, and carcinomas, such as infiltrating ductal carcinoma, infiltrating lobular carcinoma, infiltrating ductal and lobular carcinoma, medullary carcinoma, mucinous carcinoma, comedocarcinoma, Paget's Disease, papillary carcinoma, tubular carcinoma, and inflammatory carcinoma.

In further aspects, the prostate cancer is selected from the group consisting of adenocarcinomas and sarcomas, and pre-cancerous conditions, such as prostate intraepithelial neoplasia.

- A method of detecting a histone deacetylase polynucleotide comprising:

  a. incubating the isolated polynucleotide according to any one of the preceding embodiments with a biological sample under conditions to allow the polynucleotide to hybridize with a polynucleotide in the sample to form a complex; and

  b. identifying the complex formed in (a), wherein identification of the complex indicates detection of a histone deacetylase polynucleotide.

- A method of detecting a histone deacetylase polypeptide comprising:

  a. incubating the antibody according to any one of the preceding embodiments with a biological sample under conditions to allow the antibody to associate with a polypeptide in the sample to form a complex; and

  b. identifying the complex formed in (a), wherein identification of the complex indicates detection of a histone deacetylase polypeptide.

- A method of screening test agents to identify modulating agents capable of altering deacetylase activity of a histone deacetylase polypeptide comprising:

  a. contacting the isolated polypeptide according to any one of the preceding embodiments with test agents under conditions to allow

the polypeptide to associate with one or more test agents; and

b. selecting test agents that alter the deacetylase activity of the polypeptide, whereby this alteration indicates identification of modulating agents. In various aspects, the modulating agents are selected from the group consisting of antagonists and inhibitors of histone deacetylase activity. In other aspects, the modulating agents are selected from the group consisting of agonists or activators of histone deacetylase activity.

- A method for screening test agents to identify modulating agents which inhibit or antagonize deacetylation activity of a histone deacetylase, comprising:

  a. combining an isolated polypeptide according any one of the preceding embodiments having a histone deacetylase activity with a histone deacetylase substrate and a test agent in a reaction mixture; and

  b. determining the conversion of the substrate to product; wherein a statistically significant decrease in the conversion of the substrate in the presence of the test agent indicates identification of a modulating agent which inhibits or antagonizes the deacetylation activity of histone deacetylase.

- A method for screening test agents to identify modulating agents that inhibit or antagonize interaction of histone deacetylase with a histone deacetylase binding protein, comprising:

  a. combining the isolated polypeptide according any one of the preceding embodiments having a histone deacetylase activity with the histone deacetylase binding protein and a test agent in a reaction mixture; and

  b. detecting the interaction of the polypeptide with the histone deacetylase binding protein to form a complex; wherein a statistically significant decrease in the interaction of the polypeptide and protein in the presence of the test agent indicates identification of a modulating agent which inhibits or antagonizes interaction of the histone deacetylase

polypeptide with the histone deacetylase binding protein.

In various aspects, one or both of the histone deacetylase polypeptide and the histone deacetylase binding protein is a fusion protein. In other aspects, at least one of the histone deacetylase polypeptide and the histone deacetylase binding protein comprises a detectable label for detecting the formation of the complex. In a further aspect, the interaction of the histone deacetylase polypeptide and the histone deacetylase binding protein is detected in a two-hybrid assay system.

- A method of screening a library of molecules or compounds to identify at least one molecule or compound therein which specifically binds to a histone deacetylase polynucleotide, comprising:

  a. combining the isolated polynucleotide according to any one of the preceding embodiments with a library of molecules or compounds under conditions to allow specific binding of the polynucleotide to at least one of the molecules or compounds; and b. detecting the specific binding in (a), thereby identifying a molecule or compound which specifically binds to the histone deacetylase polynucleotide. In various aspects, the library comprises molecules selected from the group consisting of selected from the group consisting of DNA molecules, RNA molecules, artificial chromosomes, PNAs, peptides, and polypeptides. In one aspect, the detecting is performed by the use of high throughput screening.

- A method of treating a disease or disorder associated with abnormal cell growth or proliferation in a mammal comprising administrating the antagonist or inhibitor of histone deacetylase polypeptide according to any one of the preceding embodiments in an amount effective to treat the disease or disorder.

In various aspects, the disease or disorder is selected from neoplasms, tumors and cancers.

- A method of treating a disease or disorder associated with abnormal cell growth or proliferation in a mammal comprising administrating the antisense polynucleotide according to any one of the preceding embodiments in an amount effective to treat the disease or disorder.

In various aspects, the disease or disorder is selected from neoplasms, tumors and cancers.

- A method of modulating one or more of cell growth or proliferation, cell differentiation, or cell survival of a eukaryotic cell, comprising combining the cell with an effective amount of a modulating agent that alters the deacetylase activity of a histone deacetylase polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, and SEQ ID NO:95, and thereby modulating the rate of one or more of cell growth or proliferation, cell differentiation, or cell survival of the eukaryotic cell, relative to the effect on the eukaryotic cells in the absence of the modulating agent.

EXAMPLES

The Examples below are provided to illustrate the subject invention and are not intended to limit the invention in any way.

EXAMPLE 1: IDENTIFICATION OF NOVEL HDAC GENE FRAGMENTS

Gene fragments encoding the novel HDAC (HDAL) polypeptides of this invention were identified by a combination of the following methods. Homology-based searches using the TBLASTN program (S.F. Altschul et al., 1997, *Nucl. Acids Res.*, 25(17):3389-3402) were performed to compare known histone deacetylases with human genomic (gDNA) and EST sequences. EST or gDNA sequences having significant homology to one or more of phosphatases (expect score less than or equal to $1x10^{-3}$) were retained for further analysis.

Hidden Markov Model (HMM) searches using PFAM motifs (listed in Table 2) (A. Bateman et al., 1999, *Nucleic Acids Research*, 27:260-262 and E.L. Sonnhammer et al., 1997, *Proteins*, 28(3):405-420) to search human genomic sequence using the Genewise program. EST or gDNA sequences

having a significant score (greater than or equal to 10) with any of the following motifs were retained for further analysis.

HMM searches using PFAM motifs (listed in Table 1) to search predicted protein sequences identified by GENSCAN analysis of human genomic sequence (C. Burge and S. Karlin, 1997, *J. Mol. Biol.*, 268(1):78-94). gDNA sequences having a significant score (greater than or equal to 10) with any of the following motifs were retained for further analysis.

Table 1: PFAM motifs used to identify histone deacetylases

Motif Name	PFAM Accession #	Description
Hist_deacetyl	PF00850	Histone deacetylase family (length 342)

Once a bacterial artificial chromosome (BAC) encoding a novel histone deacetylase-like protein was identified by any of the methods listed above, its predicted protein sequence was used to identify the most closely related known histone deacetylase using the BLASTP program(NCBI). This known protein was used as the query for a GenewiseDB search of the original BAC and all nearby BACs (identified by the Golden Path tiling map, UCSC). The results were used to identify additional potential exons, intron/exon boundaries, partial transcript cDNA sequence and partial predicted protein sequence for the novel HDAC gene. The Primer3 program (S. Rozen et al., 1998, 0.6 Ed., Whitehead Institute Center for Genomic Research, Cambridge, MA) was used to design PCR primers within single exons and between adjacent exons and to design antisense 80mer probes for use in isolating cDNA clones.

EXAMPLE 2: ANALYSIS OF HDACs

Enzymatic Activity Measurements

Constructs representing the open reading frames of the identified novel sequences are engineered in frame with c-MYC or FLAG epitopes using commercially available mammalian expression vectors. These plasmids are transfected into HEK293 or COS7 cells and novel HDAC protein expression are analyzed by Western .blot analysis of protein lysates from the transfectants using anti-MYC epitope or anti-FLAG epitope antibodies.

MYC or FLAG tagged-HDAC proteins are immunoprecipitated from the lysates and incubated with {^{3}H} acetate- or fluorescent-labeled acetylated proteins. Release of {^{3}H} acetate or decrease in fluorescent signal intensity is used to establish the activity of the putative HDACs. The effects of pan-HDAC chemical inhibitors on the enzymatic activity of the novel HDACs is also assessed and compared with the activity of known HDAC proteins and their inhibition with these chemical agents.

Transcriptional Assays

HDAC proteins have been shown to positively or negative regulate transcriptional pathways. The ability of the novel HDAC proteins to repress or activate the constitutive or regulated activity of transcriptional reporter plasmids is assessed. These assays are performed using transient transfections of mammalian expression constructs encoding the novel HDAC proteins with reporter plasmid constructs of containing response elements of specific transcriptional pathways (e.g., p53, AP1, androgen receptor, LEF1/TCF4), a minimal promoter and a reporter gene product (e.g., alkaline phosphatase, luciferase, green fluorescent protein).

Alternatively, the novel HDACs are transfected into cell lines engineered to stably express these transcriptional reporter plasmids. Because the consequence of HDAC expression could be inhibitory or stimulatory, the effects of the novel HDAC proteins on these transcriptional responses are monitored in the presence and absence of activators of the pathway. Similar to enzymatic activity measurements, pan-inhibitors of the known HDACs are also examined to establish the enzymatic activity of the novel HDAC gene products as protein deacetylases.

Expression Analysis

Initial insights into the role of the novel HDACs in normal physiology and disease states is assessed by a variety of expression analyses. Quantitative reverse transcriptase polymerase chain reaction (RT-PCR) using primers specific to the novel sequences is implemented to evaluate the expression of novel HDAC mRNA in a variety of normal cell lines and tissue as well as a spectrum of human tumor cell lines. Expression profiles of novel

HDACs are confirmed using Northern blot analysis or ribonuclease protection assays.

In addition, tissue arrays containing a variety of patient organ samples and arrays of malignant tissue are evaluated by *in situ* hybridization to gain further insights into the association of the novel HDAC proteins with particular physiological responses and in neoplasia.

Subcellular Localization

The subcellular localization of MYC- or FLAG-tagged novel HDAC proteins is determined upon ectopic expression in mammalian cells. Cells are fixed, permeabilized and incubated with anti-MYC or anti-FLAG antibodies to detect expressed protein. The localization of tagged proteins is then detected using CY3 or FITC-conjugated secondary antibodies and visualized by fluorescent microscopy. These studies can determine if the assayed HDACs deacetylate nuclear or cytoplasmic protein substrates.

EXAMPLE 3: OLIGONUCLEOTIDES FOR THE ISOLATION OF HDACs

BMY_HDAL1

Based on the predicted gene structure of BMY_HDAL1, the Primer3 program designed the following PCR primers and probe oligos for isolation of cDNAs. Table 2 presents single exon primers and probes for BMY_HDAL1 cDNA isolation. Table 3 presents multiple exon primers for BMY_HDAL1 cDNA isolation. Table 4 presents BMY_HDAL1 capture oligonucleotides. As shown below in Table 5, a separately designed primer set was used to test for BMY_HDAL1 expression using a cDNA pool from human placenta and the following human tumor cell lines including Caco-2, LS174-T, MIP, HCT-116, A2780, OVCAR-3, HL60, A431, Jurkat, A549, PC3 and LnCAP cells.

BMY_HDAL2

Based on the predicted gene structure of BMY_HDAL2, the Primer3 program designed the following PCR primers and probe oligonucleotides for isolation of cDNAs. BMY_HDAL2 single exon primers and probes are shown in Table 6. Multiple exon primers for BMY-HDAL2 cDNA isolation are shown in Table 7. BMY_HDAL2 capture oligonucleotides are shown in Table 8. As shown in Table 9, a separately designed primer set was used to test for

BMY_HDAL2 expression using a cDNA pool from human placenta and the following human tumor cell lines: Caco-2, LS174-T, MIP, HCT-116, A2780, OVCAR-3, HL60, A431, Jurkat, A549, PC3 and LnCAP cells.

BMY_HDAL3

Based on the predicted gene structure of BMY_HDAL3, the Primer3 program designed the following PCR primers and probe oligonucleotides for isolation of cDNAs. For BMY_HDAL3, the following primer sets were designed from the AC002410 sequence using Primer3. Single exon primers for the novel BMY-HDAL3 isolation are shown in Table 10. Multiple exon primers for BMY_HDAL3 isolation are presented in Table 11. BMY_HDAL3 capture oligonucleotides are shown in Table 12.

Table 2

Primer Set			Left Primer			Right Primer		
Template	Set	Product Size	Start, Length	Sequence	Tm	Start, Length	Sequence	Tm
BMY_HDAL1 exon 1	1	118	16, 20	ccttgatgctgaaacaccag (SEQ ID NO:24)	59.3	133, 21	tcacatttatttagcagccca (SEQ ID NO:25)	58.3
BMY_HDAL1 exon 1	2	119	16, 20	ccttgatgctgaaacaccag (SEQ ID NO:26)	59.3	134,22	ctcacatttatttagcagccca (SEQ ID NO:27)	59.3

Table 3

Primer Set			Left Primer			Right Primer		
Template	Set	Product Size	Start, Length	Sequence	Tm	Start, Length	Sequence	Tm
BMY_HDAL1 exons 1_2	1	148	67, 20	agcatgctggacgaatacag (SEQ ID NO:28)	58.9	234, 20	ttggtgccatacaacagtga (SEQ ID NO:29)	58.5
BMY_HDAL1 exons 1_2	2	199	16, 20	ccttgatgctgaaacaccag (SEQ ID NO:30)	59.3	234, 20	ttggtgccatacaacagtga (SEQ ID NO:31)	58.5
BMY_HDAL1 exons 2_3	1	110	60, 20	tcactgttgtatggcaccaa (SEQ ID NO:32)	58.5	189, 20	ccaagtccaccacaaggtaa (SEQ ID NO:33)	58.5
BMY_HDAL1 exons 2_3	2	104	60, 20	tcactgttgtatggcaccaa (SEQ ID NO:34)	58.5	183, 20	ccaccacaaggtaatgagga (SEQ ID NO:35)	58.4

Table 4

Oligo		Capture Probe	
Template	Number	Start, Size	Sequence (ANTISENSE)
BMY_HDAL2 exon 1	1	36, 77	gtttcttgcagtcgtgaccagatactctgtattcgtccagcatgctcagggtgggtggtggaattgccacaaacgca (SEQ ID NO:36)

Table 5

HDAL Gene	5'-oligo primer sequence (5'-3')	3'-oligo primer sequence (5'-3')	Predicted Product	Product observed
HDAL1	ggaattgcctatgacccettga (SEQ ID NO:37)	tgtacttaccccaagtccaccaca (SEQ ID NO:38)	316 nt	yes

Table 6

Primer Set			Left Primer			Right Primer		
Template	Set	Pro-duct Size	Start, Length	Sequence	Tm	Start, Length	Sequence	Tm
BMY_HDAL2 exon 1	1	102	2, 20	ggacagtgacaccatttgga (SEQ ID NO:39)	59.4	103, 19	agctctcctgaggccactt (SEQ ID NO:40)	59.1
BMY_HDAL2 exon 1	2	100	2, 20	ggacagtgacaccatttgga (SEQ ID NO:41)	59.4	101, 19	ctctcctgaggccactttg (SEQ ID NO:42)	58.5
BMY_HDAL2 exon 4	NA							
BMY_HDAL2 exon 5	1	103	10, 20	gccttggagaagggtacaat (SEQ ID NO:43)	58.1	112, 23	gaaagaagtaccaacctgaatgc (SEQ ID NO:44)	59.2
BMY_HDAL2 exon 5	2	102	10, 20	gccttggagaagggtacaat (SEQ ID NO:45)	58.1	111, 22	aaagaagtaccaacctgaatgc (SEQ ID NO:46)	57.4

Table 7

Primer Set			Left Primer			Right Primer		
Template	Set	Product Size	Start, Length	Sequence	Tm	Start, Length	Sequence	Tm
BMY_HDAL2 exons 1-2	1	157	2, 20	ggacagtgacaccatttgga (SEQ ID NO:47)	59.4	178, 2	tgtggattcttcagcgtgat (SEQ ID NO:48)	59.2
BMY_HDAL2 exons 1-2	2	126	2, 20	ggacagtgacaccatttgga (SEQ ID NO:49)	59.4	147, 20	ctcacaacagcaaacccatt (SEQ ID NO:50)	58.6
BMY_HDAL2 exons 2-3	1	107	0, 20	aatgggtttgctgttgtgag (SEQ ID NO:51)	58.6	126, 20	tctctcaagtatttggcggt (SEQ ID NO:52)	57.4
BMY_HDAL2 exons 2-3	2	108	0, 20	aatgggtttgctgttgtgag (SEQ ID NO:53)	58.6	127, 20	gtctctcaagtatttggcgg (SEQ ID NO:54)	57.4
BMY_HDAL2 exons 3-4	1	130	23, 20	ttgcaattaccgccaaatac (SEQ ID NO:55)	58.6	172, 20	gaaatgtacaggatgctggg (SEQ ID NO:56)	58.0
BMY_HDAL2 exons 3-4	2	131	22, 20	gttgcaattaccgccaaata (SEQ ID NO:57)	58.561	172, 20	gaaatgtacaggatgctggg (SEQ ID NO:58)	58.019
BMY_HDAL2 exons 4-5	1	105	45, 20	cccagcatcctgtacatttc (SEQ ID NO:59)	58.019	169, 20	attgtacccttctccaaggc (SEQ ID NO:60)	58.121
BMY_HDAL2 exons 4-5	2	113	69, 20	catcgctatgatgaagggaa (SEQ ID NO:61)	58.671	201, 18	ggatcaaggccacctgtc (SEQ ID NO:62)	58.969

Table 8

Set		Capture Probe	
Template	Oligo Number	Start, Size	Sequence (ANTISENSE)
BMY_HDAL2 exon 1	No oligo		
BMY_HDAL2 exon 4	1	23, 80	tgccagggaaaaagttcccttcatcatagcgatggagtgaaatgtacaggatgctggggtcagcataaaaggcctgctg g (SEQ ID NO:63)
BMY_HDAL2 exon 4	2	19, 79	gggaaaaagttcccttcatcatagcgatggagtgaaatgtacaggatgctggggtcagcataaaaggcctgctgggtac (SEQ ID NO:64)

Table 9

HDAL Gene	5'-oligo primer sequence (5'-3')	3'-oligo primer sequence (5'-3')	Predicted Product	Product observed
HDAL2	gtggacagtgacaccatttgga (SEQ ID NO:65)	ggagaaagaagtaccaacctgaatgctt (SEQ ID NO:66)	489 nt	yes

Table 10

Primer Set			Left Primer			Right Primer		
Template	Set	Product Size	Start, Length	Sequence	Tm	Start, Length	Sequence	Tm
BMY_HDAL3 exon 1	1	100	18, 20	gtggccaaagagtttgatcc (SEQ ID NO:67)	60	117, 20	ttgccgtcactttgtaccct (SEQ ID NO:68)	60
BMY_HDAL3 exon 1	2	100	18, 20	gtggccaaagagtttgatcc (SEQ ID NO:69)	60	117, 19	ttgccgtcactttgtaccc (SEQ ID NO:70)	59
BMY_HDAL3 exon 2	1	120	4, 20	tggtcatttgacgaagcaat (SEQ ID NO:71)	59	123, 20	agaagggcatttacacaggc (SEQ ID NO:72)	59
BMY_HDAL3 exon 2	2	119	4, 20	tggtcatttgacgaagcaat (SEQ ID NO:73)	59	122, 20	gaagggcatttacacaggct (SEQ ID NO:74)	59

Table 11

Primer Set			Left Primer			Right Primer		
Template	Set	Product Size	Start, Length	Sequence	Tm	Start, Length	Sequence	Tm
BMY_HDAL3 exons 1-2	1	147	95, 20	aggagggtacaaagtgacgg (SEQ ID NO:75)	59	261, 20	agggcatttacacaggcttc (SEQ ID NO:76)	59
BMY_HDAL3 exons 1-2	2	146	95, 20	aggagggtacaaagtgacgg (SEQ ID NO:77)	59	260, 20	gggcatttacacaggcttct (SEQ ID NO:78)	59
BMY_HDAL3 exons 2-3	1	160	25, 20	gatgacattggctgatggac (SEQ ID NO:79)	59	204, 20	agcattcatattcgggcttt (SEQ ID NO:80)	59
BMY_HDAL3 exons 2-3	2	181	4, 20	tggtcatttgacgaagcaat (SEQ ID NO:81)	59	204, 20	agcattcatattcgggcttt (SEQ ID NO:82)	59

Table 12

Set		Capture Probe	
Template	Set	Start, Size	Sequence (ANTISENSE)
BMY_HDAL3 exon 1	1	32, 80	tcactttgtaccctcctagaggaggggtgtggccttccaatgcatcaaatccagcagatactaagaccatgtctggatca (SEQ ID NO:83)
BMY_HDAL3 exon 1	2	19, 80	tcctagaggaggggtgtggccttccaatgcatcaaatccagcagatactaagaccatgtctggatcaaactctttggcca (SEQ ID NO:84)
BMY_HDAL3 exon 2	1	27, 80	ggcttctgatgcatcacagatggctgtgagatcatgtcctccttctagagccaacaccacacgtccatcagccaatgtca (SEQ ID NO:85)
BMY_HDAL3 exon 2	2	27, 80	ggcttctgatgcatcacagatggctgtgagatcatgtcctccttctagagccaacaccacacgtccatcagccaatgtca (SEQ ID NO:86)

EXAMPLE 4: COMPLEMENTARY POLYNUCLEOTIDES

Antisense molecules or nucleic acid sequence complementary to an HDAC protein-encoding sequence, or any part thereof, can be used to decrease or to inhibit the expression of naturally occurring HDAC. Although the use of antisense or complementary oligonucleotides comprising about 15 to 35 base-pairs is described, essentially the same procedure is used with smaller or larger nucleic acid sequence fragments. An oligonucleotide based on the coding sequence of an HDAC polypeptide or peptide, for example, as shown in FIG. 1, FIG. 5, FIG. 10, FIGS. 15A-15C, FIGS. 20A-20C, and FIGS. 21A-21B, and as depicted in SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, or SEQ ID NO:96, for example, is used to inhibit expression of naturally occurring HDAC. The complementary oligonucleotide is typically designed from the most unique 5' sequence and is used either to inhibit transcription by preventing promoter binding to the coding sequence, or to inhibit translation by preventing the ribosome from binding to an HDAC protein-encoding transcript.

Using a portion SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, or SEQ ID NO:96, for example, an effective antisense oligonucleotide includes any of about 15-35 nucleotides spanning the region which translates into the signal or 5' coding sequence of the HDAC polypeptide. Appropriate oligonucleotides are designed using OLIGO 4.06 software and the HDAC coding sequence (e.g., SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, or SEQ ID NO:96).

EXAMPLE 5: NORTHERN BLOT ANALYSIS FOR HDACs

Northern Blot analysis is used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNA from a particular cell or tissue type has been bound (See, J. Sambrook et al., *supra*). Analogous computer techniques using BLAST (S.F. Altschul, 1993, *J. Mol. Evol.*, 36:290-300 and S.F. Altschul et al., 1990, *J. Mol. Evol.*, 215:403-410) are used to search for identical or related molecules in nucleotide databases, such as GenBank or the LIFESEQ database (Incyte Pharmaceuticals). This analysis is much more rapid and

less labor-intensive than performing multiple, membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as being exact (identical) or homologous.

The basis of the search is the product score, which is defined as follows: (% sequence identity x maximum BLAST score) / 100. The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1-2% error; at 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules. The results of Northern analysis are reported as a list of libraries in which the transcript encoding HDAC polypeptides occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times that a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences that are examined in the cDNA library.

EXAMPLE 6: MICROARRAYS FOR ANALYSIS OF HDACs

For the production of oligonucleotides for a microarray, an HDAC sequence, e.g., a novel HDAC having SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, or SEQ ID NO:96, for example, is examined using a computer algorithm which starts at the 3' end of the nucleotide sequence. The algorithm identifies oligomers of defined length that are unique to the gene, have a GC content within a range that is suitable for hybridization and lack predicted secondary structure that would interfere with hybridization. The algorithm identifies specific oligonucleotides of 20 nucleotides in length, i.e., 20-mers. A matched set of oligonucleotides is created in which one nucleotide in the center of each sequence is altered. This process is repeated for each gene in the microarray, and double sets of 20-mers are synthesized in the presence of fluorescent or radioactive nucleotides and arranged on the surface of a substrate. When the substrate

is a silicon chip, a light-directed chemical process is used for deposition (WO 95/11995, M. Chee et al.).

Alternatively, a chemical coupling procedure and an ink jet device is used to synthesize oligomers on the surface of a substrate. (WO 95/25116, J.D. Baldeschweiler et al.). As another alternative, a "gridded" array that is analogous to a dot (or slot) blot is used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using, for example, a vacuum system, or thermal, UV, mechanical, or chemical bonding techniques. A typical array may be produced by hand, or by using available materials and equipment, and may contain grids of 8 dots, 24 dots, 96 dots, 384 dots, 1536 dots, or 6144 dots. After hybridization, the microarray is washed to remove any non-hybridized probe, and a detection device is used to determine the levels and patterns of radioactivity or fluorescence. The detection device may be as simple as X-ray film, or as complicated as a light scanning apparatus. Scanned fluorescent images are examined to determine degree of complementarity and the relative abundance/expression level of each oligonucleotide sequence in the microarray.

EXAMPLE 7: PURIFICATION OF HDAC POLYPEPTIDES

Naturally occurring or recombinant HDAC polypeptide is substantially purified by immunoaffinity chromatography using antibodies specific for an HDAC polypeptide, or a peptide derived therefrom. An immunoaffinity column is constructed by covalently coupling anti-HDAC polypeptide antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Medium containing HDAC polypeptide is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of the HDAC polypeptide (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/HDAC polypeptide binding (e.g., a buffer of pH 2-3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and HDAC polypeptide is collected.

EXAMPLE 8: IDENTIFICATION OF MOLECULES THAT INTERACT WITH HDAC POLYPEPTIDES

HDAC polypeptides, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton et al., 1973, *Biochem. J.*, 133:529). Candidate molecules previously arrayed in wells of a multi-welled plate are incubated with the labeled HDAC polypeptide, washed, and any wells having labeled HDAC polypeptide-candidate molecule complexes are assayed. Data obtained using different concentrations of HDAC polypeptide are used to calculate values for the number, affinity and association of an HDAC polypeptide with the candidate molecules.

Another method suitable for identifying proteins, peptides or other molecules that interact with an HDAC polypeptide include ligand binding assays such as the yeast-two hybrid system as described hereinabove.

EXAMPLE 9: IDENTIFICATION AND CLONING OF HDAC9c

Bioinformatic searches of the assembled human genome sequence were performed using a conserved consensus sequence derived from the catalytic domain of class I and class II HDACs. Three gene fragments (HDAL1, HDAL2, HDAL3) were identified from the assembled sequence of human chromosome 7q36 that encoded amino acids sequence with homology to class II HDACs. Biotinylated single stranded oligonucleotides representing unique sequences from these predicted gene fragments of the following sequence were prepared:

HDAL1, 5-gtttcttgcagtcgtgaccagatactctgattcgtccagcatgctcagggt gggtgggtggaattgccacaaacgca (SEQ ID NO:101);

HDAL2, 5'-tgccagggaaaaagt tcccttcatcatagcgatggagtgaaatgtaca ggatgctggggtcagcataaaggcctgctgg (SEQ ID NO:102); and

HDAL3, 5' tgatccagacatggtcttagtatctgctggatttgatgcattggaaggcca cacccctcctctaggagggtacaaagtga (SEQ ID NO:103).

The biotinylated oligonucleotides were hybridized to fractions of cDNA prepared from human placenta, and positive sequences were identified by PCR. Three of the clones identified (HDACX1A, HDACX2A, and HDACX3A) contained overlapping cDNAs that showed sequence identity to the predicted

gene fragments. These cDNAs encoded a novel sequence, designated HDAC9c (FIGS. 15A-15C), that shared homology to class II HDACs. A full length HDAC9c construct was prepared by combining a 1.3 kb *Bam*HI-*Pst*I fragment from the HDACX2A clone with a 3.5 kb *Pst*I-*Not*I fragment from the HDACX3A. These fragments were ligated into mammalian expression vectors pcDNA3.1 and pcDNA4.0. The resulting constructs were evaluated by DNA sequencing to confirm the identity of the inserts. The HDAC9c pcDNA3.1 construct was deposited at the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209 on June 12, 2002 under ATCC Accession No. ____________________ according to the terms of the Budapest Treaty.

Three fragments that encoded homology to class II HDACs were identified from the assembled sequence of human chromosome 7q36. Subsequent cDNA cloning bioinformatics analysis revealed that these gene fragments encoded a single class II HDAC, comprising a protein of 1147 amino acids. This sequence was provisionally designated as HDAC-9, and later renamed HDAC9c. During the course of this work, similar sequences were reported by Zhou et al. (2001, *Proc. Natl. Acad. Sci. USA* 98:10572-7), including two isoforms related to class II HDAC proteins. Sequence alignments revealed the HDAC-9 sequence was closely related to the previously identified HDAC9 sequences (GenBank Accession Nos. AY032737 and AY032738). However, the published sequences lacked a large portion of the C-terminal domain common to known class HDAC proteins (FIGS. 15D-15F).

One of the HDAC9 isoforms (HDAC9a, (GenBank Accession No. AY032737) lacked ~ 185 C-terminal amino acids compared to other HDAC family members. Another isoform of HDAC9 (HDAC9, (GenBank Accession No. AY032738) lacked approximately 65 C-terminal amino acids compared to other HDAC family members. In contrast to these sequences, the HDAC9c sequence, also designated as HDAC-X, contained more than 50 additional amino acids at its C-terminus (FIGS. 15D-15F). The HDAC9c sequence was deemed to represent the full-length version of HDAC9. Notably, HDAC9c

contained an LQQ sequence motif at positions 123-125. This motif was missing in the HDAC9 C-terminal truncated isoforms, but was conserved in other HDAC family members. Thus, the LQQ sequence motif may be important for the function of the HDAC9c protein. No other motifs were identified by PFAM analysis (A. Bateman et al., 2002, *Nucl. Acids Res.* 30:276-80).

EXAMPLE 10: EXPRESSION PROFILING FOR HDAC9

To determine the distribution of HDAC9 in adult normal tissues, the expression profile of HDAC9 was examined by Northern blot analysis. Northern blotting was performed as described (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edition). Tissue samples were obtained from CLONTECH (Palo Alto, CA). The probe for Northern blotting was derived from nucleotides 2917-3211 of HDAC9c (FIG. 16D; SEQ ID NO:92). Two > 8.0 kb HDAC9 transcripts were detected at low levels in brain, skeletal muscle, stomach, and trachea tissue (FIG. 16A). Upon longer exposure, HDAC9 mRNA was also detected in mammary gland and prostate tissue (FIG. 16A).

Given the low level of expression in normal tissues, experiments were performed to determine the expression of HDAC9 in human tumor cell lines. HDAC9 mRNA expression levels were evaluated by quantitative PCR analysis on first-strand cDNA prepared from a variety of human tumor cell lines (ATCC, Rockville, MD). HDAC9 levels were normalized to GAPDH mRNA levels within the samples, and RNA levels were quantified using the fluorophore SYBR green. For amplification, HDAC9 primers were used: forward primer 5'-gtgacaccatttggaatgagctac (SEQ ID NO:104); and reverse primer 5'ttggaagccagctcgatgac (SEQ ID NO:105). HDAC9 expression was found to be elevated in ovarian, breast, and certain lung cancer cell lines (FIG. 16B). In contrast, HDAC9 was poorly expressed in tumor cell lines derived from colon tumor specimens (FIG. 16B).

To confirm these results, nuclease protection experiments were performed on RNAs isolated from select tumor cell displaying a range of HDAC9 expression. Nuclease protection was performed using ^{35}S-labeled

UTP as a radioactive precursor for a in accordance with published methods (Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edition). The riboprobe sequence was derived from nucleotides 2917-3211 in HDAC9c (FIG. 16D; SEQ ID NO:92). Brain tissue was included as a control to show normal tissue expression levels. The profile of HDAC9 expression observed by quantitative RT-PCR was confirmed by nuclease protection (i.e., A2780 > MDA-MB453 > MCF7; FIG. 16C). The pervasive expression of HDAC9 in tumor cell lines of diverse origin, and the low level expression of HDAC9 in normal adult tissue, suggested that the expression of this gene was regulated in tumor progression.

<u>EXAMPLE 11: *IN SITU* HYBRIDIZATION TO ANALYZE HDAC9 EXPRESSION</u>

To further analyze the upregulation of HDAC9 in tumor cells, a variety of human tumor and normal tissue specimens were subjected to *in situ* hybridization using an HDAC9 antisense riboprobe and tissue microarrays. A ^{35}S-labeled cRNA riboprobe was prepared from a 295 bp cDNA fragment from the HDAC9 coding region (FIG. 16D; SEQ ID NO:92). This fragment encoded the most divergent region of the HDAC9 protein. The riboprobe was hybridized to paraffin-embedded clinical tissue specimens derived from normal or cancerous tissues, and processed by standard procedures (Lorenzi et al., 1999, *Oncogene* 18:4742-4755). Hybridized sections were incubated for 3 to 6 weeks, and the level and localization of HDAC9 staining was evaluated by microscopy. Staining levels were quantified by a board-certified pathologist.

HDAC9 mRNA levels were generally below the limit of detection (staining level = 0) in normal tissues, including breast, kidney, testis, and liver tissues. Low to moderate levels of HDAC9 mRNA (staining level = 1-2) were detected in lymph node, brain, adrenal gland, pancreas, bladder, lung, and gastric tissues (data not shown). Normal breast and prostate tissue showed average staining levels of 0 and 1, respectively (FIGS. 17A-17C). A dramatic increase in HDAC9 mRNA expression was detected in breast tumor (average staining level = 2-3) and prostate tumor (average staining level = 2) tissues

(FIGS. 17A-17C). Preliminary data also showed increased expression of HDAC9 in endometrial and ovarian tumors. Thus, HDAC9 was expressed at very low levels in normal adult peripheral tissues, but was overexpressed in a variety of tumors, including breast and prostate adenocarcinomas. This suggested that HDAC9 expression correlated with the progression of breast and prostate tumors.

EXAMPLE 12: EFFECT OF HDAC9c ON CELLULAR TRANSFORMATION

Results of the experiments, above, indicated that elevated HDAC9c expression was associated with certain tumor cells. To further investigate its involvement in tumorogenesis, HDAC9c was evaluated for its ability to morphologically transform mouse fibroblasts. HDAC9c in pcDNA3.1 was introduced by calcium phosphate transfection into 1.5 x 10^5 NIH/3T3 cells (ATCC, Rockville, MD) in duplicate at 1.0 μg/10 cm plate. One set of cultures received growth medium (DMEM containing 5% calf serum) while the parallel culture received growth medium containing 750 μg/ml of G418 to develop stable clonal populations.

After 10-14 days in culture, unselected plates were stained with Geimsa (Sigma-Aldrich, St. Louis, MO), and morphologically transformed foci were visualized. Selected clones were examined for growth in soft agar at 10^5, 10^4, or 10^3 cells/15 mm well following standard protocols. After 2-3 weeks in culture, colonies were visualized by microscopy and tetrazolium violet staining. HDAC9c transfectants produced some foci in monolayer culture (data not shown). However, the response was not robust, suggesting that higher levels HDAC9c expression levels were required to transform NIH/3T3 cells.

HDAC9c transfectants were also evaluated for anchorage-independent growth. NIH/3T3 cells stably transfected with HDAC9c or FGF8 constructs, or vector alone, were suspended in soft agar containing growth medium and cultured for 2-3 weeks. FGF8 is a cDNA that potently transforms NIH/3T3 through autocrine stimulation of endogenous FGF receptors (Lorenzi et al., 1995, Oncogene 10:2051-2055). In vector transfectants, very few colonies greater than 50 μm in diameter were observed after three weeks in culture

(FIG. 18). In contrast, FGF8 transfectants produced several colonies greater than 50 µ□m after three weeks (FIG. 18). HDAC9c transfectants also produced significant colony growth compared to vector transfectants, but less than that observed for FGF8 transfectants (FIG. 18). These results suggested that overexpression of HDAC9c induced an oncogenic phenotype in mouse fibroblasts.

EXAMPLE 13: EFFECT OF HDAC9c ON THE ACTIN CYTOSKELETON

Changes in the actin cytoskeleton often accompany the transformed phenotype of cells expressing oncogenes such as Ras, Rho, or src. In general, gene products that affect cell adhesion or motility are associated with changes in the actin cytoskeleton. To investigate whether the transformation induced by HDAC9c was associated with changes in the cytoskeletal architecture, NIH/3T3 transfectants expressing HDAC9c were subjected to fluorescent staining with TRITC-conjugated phalloidin to visualize filamentous actin (F-actin).

In these experiments, a HDAC4 construct was used as a control. For the control construct, full-length HDAC4 cDNA was amplified by RT-PCR from first-strand cDNA based on the sequence reported by Grozinger et al. (*Proc. Natl. Acad. Sci. USA* 96:4868-4873), and cloned into pcDNA3.1. Mass-selected stable NIH/3T3 clones of HDAC9c (in pcDNA3.1), Ras, HDAC4, or vector alone, were plated in 8 well chamber slides in duplicate and allowed to adhere overnight in growth medium (DMEM high glucose containing 10% calf serum). Cells were subsequently serum-starved for 18 hours and one set was stimulated with 10% calf serum for 15 minutes. The cultures were fixed for 30 minutes in 4% paraformaldehyde, permeabilized in 0.02% Triton-X100, and incubated with TRITC or FITC conjugated phalloidin (Sigma, St. Louis, MO) for 2 hours. Filamentous actin was visualized by fluorescence microscopy, and images were captured with a digital camera.

In parental NIH/3T3 cells (data not shown) or vector transfectants, low levels of F-actin stress fiber formation were observed following serum starvation for 18 hours (FIG. 19). Stimulation of these cells for 15 minutes with serum promoted an extensive stress fiber network (FIG. 19), indicating

that the extracellular signals regulating these pathways were intact in these cells. A dramatic increase in stress fiber content and organization was observed in serum starved HDAC9c-expressing cells (FIG. 19), indicating that that expression of HDAC9c was sufficient to induce reorganization of the actin cytoskeleton. In contrast, no stress fiber formation was observed in serum starved NIH/3T3 cells expressing the HDAC4 protein (FIG. 19). These results suggested that induction of actin stress fiber formation underlay the transformed phenotype associated with expression of HDAC9c.

Conclusion

Inhibitors of HDAC activity are involved in the regulation of cellular proliferation, apoptosis, and differentiation of a variety of cell types. However, little is known about the role of individual HDACs in tumor cells or in their genesis. In accordance with the present invention, a unique HDAC isoform, HDAC9c, has been identified and characterized. HDAC9 shows restricted expression in normal adult tissues, but is overexpressed in several primary human tumors, including those derived from breast and prostate cancers. The overexpression of HDAC9c in *in vitro* models promoted the oncogenic transformation of fibroblasts and this transformed phenotype was associated with the induction of actin cytoskeletal stress fiber formation. These results suggest a functional consequence of HDAC9c overexpression is the promotion and/or maintenance of the transformation state of certain tumor cells.

Members of the HDAC protein family have been shown to possess potent ability to repress transcription. For instance, tumor suppressor genes p21 and gelsolin are expressed upon HDAC inhibition (Sowa et al., 1999, *Cancer Res.* 59(17):4266-70; Saito et al., 1999, *Proc. Natl. Acad. Sci. USA* 96:4592-4597). It is interesting to note that gelsolin negatively regulates the formation of the actin cytoskeleton (Sun et al., 1999, *J. Biol. Chem.* 274:33179-33182). In contrast, actin cytoskeleton formation is positively regulated by HDAC9c expression (FIG. 19). Thus, HDAC9c inhibition or overexpression may regulate gelsolin levels, and this regulation may underlie the cytoskeletal changes mediated by HDAC9c.

HDAC9 was overexpressed greater than 90% of the breast and prostate tumor specimens examined compared to corresponding tissue from normal patients (FIGS. 17A-17B). By comparison, the epidermal growth factor (EGF) receptor, erbB2, has been estimated to be overexpressed in roughly 30% of certain tumor types (King et al., 1985, *Science* 229:974-976). These observations strongly suggest that HDAC9c can be used as a diagnostic marker for breast or prostate tumorigenesis. Hormonal signaling is critical to the progression and treatment of breast cancers, and HDAC9 has been implicated in transcription (Zhou et al., *Proc. Natl. Acad. Sci. USA* 98:10572-10577). Without wishing to be bound by theory, it is possible that HDAC9 regulates estrogen or androgen responsive promoters in these tumor cells. As shown herein, HDAC9 expression is increased in primary cancers, and restricted in normal tissue expression. Further, HDAC9c expression induces oncogenic transformation. The sum of these observations indicates that HDAC9c can be used as a diagnostic and/or therapeutic target for certain tumors or cancers, in particular, breast and prostate tumors or cancers.

EXAMPLE 14: HDAC9 SPLICE VARIANTS

Using the methods described herein, HDAC9 splice variants were identified, including BMY_HDACX variant 1 (FIGS. 20A-20C; SEQ ID NO:94; also called BMY_HDACX_v1 and HDACX_v1) and BMY_HDACX variant 2 (FIGS. 21A-21B; SEQ ID NO:96; also called BMY_HDACX_v2 and HDACX_v2). The cDNA sequences for BMY_HDACX_v1 (SEQ ID NO:94) and BMY_HDACX_v2 (SEQ ID NO:96) were aligned to the nucleotide sequences of three reported splice products of the HDAC9 gene, including HDAC9v1 (NCBI Ref. Seq. NM_058176; FIGS. 22A-22C; SEQ ID NO:97), HDAC9v2 (NCBI Ref. Seq. NM_058177; FIGS. 22D-22F; SEQ ID NO:98), and HDAC9v3 (NCBI Ref. Seq. NM_014707; FIGS. 22G-22I; SEQ ID NO:100). The sequence alignment produced by ClustalW (D.G. Higgins et al., 1996, *Methods Enzymol.* 266:383-402) is shown in FIGS. 23A-23K.

ClustalW sequence alignments indicated that the HDAC9c amino acid sequence showed 80.5% identity to the HDAC9a (AY032738) amino acid sequence, 94.1% identity to the HDAC9 (AY032737) amino acid sequence,

and 55.1% identity to the HDAC5 (AF132608) amino acid sequence. The HDAC9c nucleotide sequence showed 81.4% identity to the HDAC9a (AY032738) nucleotide sequence, 94.3% identity to the HDAC9 (AY032737) nucleotide sequence, and 60.1% identity to the HDAC5 (AF132608) nucleotide sequence. In addition, the HDACX_v2 amino acid sequence showed 55.2% identity to the most closely related amino acid sequence, and the HDACX_v2 nucleotide sequence showed 55.3% identity to the HDAC9a (AY032738) nucleotide sequence, 48.1% identity to the HDAC9 (AY032737) nucleotide sequence, and 27.6% identity to the HDAC5 (AF132608) nucleotide sequence.

Additional amino acid sequence alignments are shown in FIGS. 24A-24D and FIGS. 25A-25C. For reference, the SEQ ID NOs of the sequences of the present invention are listed in the table shown below. HDACX_v1 and HDACX_v2 constructs were deposited at the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209 on ____________ under ATCC Accession No. ____________ ______ according to the terms of the Budapest Treaty.

Description	SEQ ID NO:
BMY_HDAL1 nucleic acid sequence	SEQ ID NO:1
BMY_HDAL1 amino acid sequence	SEQ ID NO:2
BMY_HDAL1 reverse nucleic acid sequence	SEQ ID NO:3
BMY_HDAL2 amino acid sequence	SEQ ID NO:4
BMY_HDAL3 amino acid sequence	SEQ ID NO:5
SC_HDA1 amino acid sequence	SEQ ID NO:6
Human HDAC4 amino acid sequence	SEQ ID NO:7
Human HDAC5 amino acid sequence	SEQ ID NO:8
Human HDAC7 amino acid sequence	SEQ ID NO:9
Aquifex ACUC HDAL amino acid sequence	SEQ ID NO:10
AC002088 nucleic acid sequence	SEQ ID NO:11
BMY_HDAL2 nucleic acid sequence	SEQ ID NO:12
BMY_HDAL2 reverse nucleic acid sequence	SEQ ID NO:13
AC002410 nucleic acid sequence	SEQ ID NO:14

Description	SEQ ID NO:
N-terminus of BMY_HDAL3	SEQ ID NO:15
C-terminus of BMY_HDAL3	SEQ ID NO:16
BAC AC004994 nucleic acid sequence	SEQ ID NO:17
BAC AC004744 nucleic acid sequence	SEQ ID NO:18
BMY_HDAL3 nucleic acid sequence	SEQ ID NO:19
BMY_HDAL3 reverse strand nucleic acid sequence	SEQ ID NO:20
AAC78618 amino acid sequence	SEQ ID NO:21
AAD15364 amino acid sequence	SEQ ID NO:22
AA287983 nucleic acid sequence	SEQ ID NO:23
BMY_HDAL1 single exon primer	SEQ ID NO:24
BMY_HDAL1 single exon primer	SEQ ID NO:25
BMY_HDAL1 single exon primer	SEQ ID NO:26
BMY_HDAL1 single exon primer	SEQ ID NO:27
BMY_HDAL1 multiple exon primer	SEQ ID NO:28
BMY_HDAL1 multiple exon primer	SEQ ID NO:29
BMY_HDAL1 multiple exon primer	SEQ ID NO:30
BMY_HDAL1 multiple exon primer	SEQ ID NO:31
BMY_HDAL1 multiple exon primer	SEQ ID NO:32
BMY_HDAL1 multiple exon primer	SEQ ID NO:33
BMY_HDAL1 multiple exon primer	SEQ ID NO:34
BMY_HDAL1 multiple exon primer	SEQ ID NO:35
BMY_HDAL1 capture oligonucleotide	SEQ ID NO:36
BMY_HDAL1 5' oligo primer	SEQ ID NO:37
BMY_HDAL1 3' oligo primer	SEQ ID NO:38
BMY_HDAL2 single exon primer	SEQ ID NO:39
BMY_HDAL2 single exon primer	SEQ ID NO:40
BMY_HDAL2 single exon primer	SEQ ID NO:41
BMY_HDAL2 single exon primer	SEQ ID NO:42
BMY_HDAL2 single exon primer	SEQ ID NO:43
BMY_HDAL2 single exon primer	SEQ ID NO:44
BMY_HDAL2 single exon primer	SEQ ID NO:45
BMY_HDAL2 single exon primer	SEQ ID NO:46
BMY_HDAL2 multiple exon primer	SEQ ID NO:47

Description	SEQ ID NO:
BMY_HDAL2 multiple exon primer	SEQ ID NO:48
BMY_HDAL2 multiple exon primer	SEQ ID NO:49
BMY_HDAL2 multiple exon primer	SEQ ID NO:50
BMY_HDAL2 multiple exon primer	SEQ ID NO:51
BMY_HDAL2 multiple exon primer	SEQ ID NO:52
BMY_HDAL2 multiple exon primer	SEQ ID NO:53
BMY_HDAL2 multiple exon primer	SEQ ID NO:54
BMY_HDAL2 multiple exon primer	SEQ ID NO:55
BMY_HDAL2 multiple exon primer	SEQ ID NO:56
BMY_HDAL2 multiple exon primer	SEQ ID NO:57
BMY_HDAL2 multiple exon primer	SEQ ID NO:58
BMY_HDAL2 multiple exon primer	SEQ ID NO:59
BMY_HDAL2 multiple exon primer	SEQ ID NO:60
BMY_HDAL2 multiple exon primer	SEQ ID NO:61
BMY_HDAL2 multiple exon primer	SEQ ID NO:62
BMY_HDAL2 capture oligonucleotide	SEQ ID NO:63
BMY_HDAL2 capture oligonucleotide	SEQ ID NO:64
BMY_HDAL2 5' oligo primer	SEQ ID NO:65
BMY_HDAL2 3' oligo primer	SEQ ID NO:66
BMY_HDAL3 single exon primer	SEQ ID NO:67
BMY_HDAL3 single exon primer	SEQ ID NO:68
BMY_HDAL3 single exon primer	SEQ ID NO:69
BMY_HDAL3 single exon primer	SEQ ID NO:70
BMY_HDAL3 single exon primer	SEQ ID NO:71
BMY_HDAL3 single exon primer	SEQ ID NO:72
BMY_HDAL3 single exon primer	SEQ ID NO:73
BMY_HDAL3 single exon primer	SEQ ID NO:74
BMY_HDAL3 multiple exon primer	SEQ ID NO:75
BMY_HDAL3 multiple exon primer	SEQ ID NO:76
BMY_HDAL3 multiple exon primer	SEQ ID NO:77
BMY_HDAL3 multiple exon primer	SEQ ID NO:78
BMY_HDAL3 multiple exon primer	SEQ ID NO:79
BMY_HDAL3 multiple exon primer	SEQ ID NO:80

Description	SEQ ID NO:
BMY_HDAL3 multiple exon primer	SEQ ID NO:81
BMY_HDAL3 multiple exon primer	SEQ ID NO:82
BMY_HDAL3 capture oligo	SEQ ID NO:83
BMY_HDAL3 capture oligo	SEQ ID NO:84
BMY_HDAL3 capture oligo	SEQ ID NO:85
BMY_HDAL3 capture oligo	SEQ ID NO:86
HDAC9c amino acid sequence	SEQ ID NO:87
HDAC9c nucleotide sequence	SEQ ID NO:88
HDAC9 (AY032737) amino acid sequence	SEQ ID NO:89
HDAC9a (AY032738) amino acid sequence	SEQ ID NO:90
HDAC4 (ALF132608) amino acid sequence	SEQ ID NO:91
HDAC9 probe	SEQ ID NO:92
BMY_HDACX_v1 amino acid sequence	SEQ ID NO:93
BMY_HDACX_v1 nucleotide sequence	SEQ ID NO:94
BMY_HDACX_v2 amino acid sequence	SEQ ID NO:95
BMY_HDACX_v2 nucleotide sequence	SEQ ID NO:96
HDAC9v1 (NM_058176) amino acid sequence	SEQ ID NO:89
HDAC9v1 (NM_058176) nucleotide sequence	SEQ ID NO:97
HDAC9v2 (NM_058177) amino acid sequence	SEQ ID NO:90
HDAC9v2 (NM_058177) nucleotide sequence	SEQ ID NO:98
HDAC9v3 (NM_014707) amino acid sequence	SEQ ID NO:99
HDAC9v3 (NM_014707) nucleotide sequence	SEQ ID NO:100
HDAL1 primer	SEQ ID NO:101
HDAL2 primer	SEQ ID NO:102
HDAL3 primer	SEQ ID NO:103
HDAC9 forward primer	SEQ ID NO:104
HDAC9 reverse primer	SEQ ID NO:105
HDAC consensus nucleotide sequence	SEQ ID NO:106
HDAC consensus amino acid sequence	SEQ ID NO:107

The contents of all patents, patent applications, published PCT applications and articles, books, references, reference manuals and abstracts

cited herein are hereby incorporated by reference in their entirety to more fully describe the state of the art to which the invention pertains.

As various changes can be made in the above-described subject matter without departing from the scope and spirit of the present invention, it is intended that all subject matter contained in the above description, or defined in the appended claims, be interpreted as descriptive and illustrative of the present invention. Many modifications and variations of the present invention are possible in light of the above teachings.

WHAT IS CLAIMED IS:

1. An isolated polynucleotide encoding a histone deacetylase polypeptide which consists of an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, and SEQ ID NO:95.

2. An isolated polynucleotide consisting of a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:12, SEQ ID NO:19, SEQ ID NO:88, SEQ ID NO:94, and SEQ ID NO:96.

3. An primer consisting of a nucleotide sequence selected from the group consisting of SEQ ID NO:24-27, SEQ ID NO:28-35, SEQ ID NO:39-46, SEQ ID NO:47-62, SEQ ID NO:65-66, SEQ ID NO:67-74, SEQ ID NO:75-82, and SEQ ID NO:104-105.

4. A probe consisting of a nucleotide sequence selected from the group consisting of SEQ ID NO:36, SEQ ID NO:63-64, SEQ ID NO:83-86, SEQ ID NO92, and SEQ ID NO:101-103.

5. A cell line comprising the isolated polynucleotide according to claim 1.

6. An expression vector comprising the isolated polynucleotide according to claim 1.

7. A host cell comprising the expression vector according to claim 6, wherein the host cell is selected from the group consisting of bacterial, yeast, insect, mammalian, and human cells.

8. An isolated polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, and SEQ ID NO:95.

9. An antibody which binds specifically to the isolated polypeptide according to claim 8, wherein the antibody is selected from the group consisting of polyclonal and monoclonal antibodies.

10. An antisense polynucleotide which consists of a nucleotide sequence selected from the group consisting of SEQ ID NO:36, SEQ ID NO:63-64, and SEQ ID NO:83-86.

11. An expression vector comprising the antisense polynucleotide according to claim 10.

12. A pharmaceutical composition selected from the group consisting of:

a. a pharmaceutical composition comprising a monoclonal antibody that specifically binds to an isolated polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, and SEQ ID NO:95, and a physiologically acceptable carrier, diluent, or excipient;

b. a pharmaceutical composition comprising an antisense polynucleotide which consists of a nucleotide sequence selected from the group consisting of SEQ ID NO:36, SEQ ID NO:63-64, and SEQ ID NO:83-86, and a physiologically acceptable carrier, diluent, or excipient; and

c. a pharmaceutical composition comprising an expression vector comprising an isolated polynucleotide encoding a histone deacetylase polypeptide which consists of an amino acid sequence selected from the group of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:87, SEQ ID NO:93, and SEQ ID NO:95, and a physiologically acceptable carrier, diluent, or excipient.

13. A method of treating a cancer selected from the group consisting of breast and prostate cancer comprising administering the pharmaceutical composition according to claim 12 in an amount effective for treating the cancer.

14. A method of diagnosing a cancer selected from the group consisting of breast and prostate cancer comprising:

a. incubating the primer according to claim 3 with a biological sample under conditions to allow the primer to amplify a polynucleotide in the sample to produce a amplification product; and

b. measuring levels of amplification product formed in (a), wherein an alteration in these levels compared to standard levels indicates diagnosis of the cancer.

15. A method of diagnosing a cancer selected from the group consisting of breast and prostate cancer comprising:

a. incubating the probe according to claim 4 with a biological sample under conditions to allow the probe to hybridize with a polynucleotide in the sample to form a complex; and b.

measuring levels of hybridization complex formed in (a), wherein an alteration in these levels compared to standard levels indicates diagnosis of the cancer.

16. A method of diagnosing a cancer selected from the group consisting of breast and prostate cancer comprising:

a. contacting the antibody according to claim 9 with a biological sample under conditions to allow the antibody to associate with a polypeptide in the sample to form a complex; and

b. measuring levels of complex formed in (a), wherein an alteration in these levels compared to standard levels indicates diagnosis of the cancer.

17. A method of detecting a histone deacetylase polynucleotide comprising:

a. incubating the probe according to claim 4 with a biological sample under conditions to allow the probe to hybridize with a polynucleotide in the sample to form a complex; and b.

identifying the complex formed in (a), wherein identification of the complex indicates detection of a histone deacetylase polynucleotide.

18. A method of detecting a histone deacetylase polypeptide comprising:

a. incubating the antibody according to claim 9 with a biological sample under conditions to allow the antibody to associate with a polypeptide in the sample to form a complex; and

b. identifying the complex formed in (a), wherein identification of the complex indicates detection of a histone deacetylase polypeptide.

19. A method of screening test agents to identify a candidate bioactive agent comprising:

a. contacting the isolated polynucleotide according to claim 1 with test agents under conditions to allow a test agent to associate with the polynucleotide to form a complex; b.

detecting the complex of (b), wherein detection of the complex indicates identification of a candidate bioactive agent.

20. A method of screening test agents to identify a candidate bioactive agent comprising:

a. contacting the isolated polypeptide according to claim 8 with test agents under conditions to allow a test agent to associate with the polypeptide to form a complex;

b. detecting the complex of (b), wherein detection of the complex indicates identification a candidate bioactive agent.

```
    GlyIleAlaTyrAspProLeuMetLeuLysHisGlnCysValCysGly
1   ggaattgcctatgaccccttgatgctgaaacaccagtgcgtttgtggc
    ccttaacggatactggggaactacgactttgtggtcacgcaaacaccg

    AsnSerThrThrHisProGluHisAlaGlyArgIleGlnSerIleTrp
49  aattccaccacccaccctgagcatgctggacgaatacagagtatctgg
    ttaaggtggtgggtgggactcgtacgacctgcttatgtctcatagacc

    SerArgLeuGlnGluThrGlyLeuLeuAsnLysCysGluArgIleGln
97  tcacgactgcaagaaactgggctgctaaataaatgtgagcgaattcaa
    agtgctgacgttctttgacccgacgatttatttacactcgcttaagtt

    GlyArgLysAlaSerLeuGluGluIleGlnLeuValHisSerGluHis
145 ggtcgaaaagccagcctggaggaaatacagcttgttcattctgaacat
    ccagcttttcggtcggacctcctttatgtcgaacaagtaagacttgta

    HisSerLeuLeuTyrGlyThrAsnProLeuAspGlyGlnLysLeuAsp
193 cactcactgttgtatggcaccaacccсctggacggacagaagctggac
    gtgagtgacaacataccgtggttggggacctgcctgtcttcgacctg

    ProArgIleLeuLeuGlyAspAspSerGlnLysPhePheSerSerLeu
241 cccaggatactcctaggtgatgactctcaaaagtttttttcctcatta
    gggtcctatgaggatccactactgagagttttcaaaaaaaggagtaat

    ProCysGlyGlyLeuGlyValSerThr
289 ccttgtggtggacttggggtaagtaca
    ggaacaccacctgaaccccattcatgt
```

FIG. 1

```
                         701                                                750
 AQUIFEX_HDAL    (12)  YGKYRYPKNHPLKIPRVSLLLRFKDAMNLIDEKELAKSRPATKEELLLFH
    BMY_HDAL1    (16)  G----NSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKASLEEIQLVH
    BMY_HDAL2     (1)  --------------------------------------------------
    BMY_HDAL3     (1)  --------------------------------------------------
         HDA4   (670)  G----SSSSHPEHAGRIQSIWSRLQETGLRGKCECIRGRKATLEELQTVH
         HDA5   (699)  G----NTHVHPEHAGRIQSIWSRLQETGLLSKCERIRGRKATLDEIQTVH
         HDA7   (496)  G----DNSRHPEHAGRIQSIWSRLQERGLRSQCECLRGRKASLEELQSVH
      SC_HDA1    (74)  TSYFEYIDPHPEDPRRIYRIYKILAENGLIN-----DPTLSGVDDLGDLM

                         751                                                800
 AQUIFEX_HDAL    (62)  TEDYINTLMEAERCQCVPKG--------------------AREKYNIGGY
    BMY_HDAL1    (62)  SEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLPCGGLGVST------
    BMY_HDAL2     (1)  -----------------------------------------VDSDTIWNE
    BMY_HDAL3     (1)  --------------------------------------------------
         HDA4   (716)  SEAHTLLYGTNPLNRQKLDSKKLLG-SLASVFVRLPCGGVGVDSDTIWNE
         HDA5   (745)  SEYHTLLYGTSPLNRQKLDSKKLLGPISQKMYAVLPCGGIGVDSDTVWNE
         HDA7   (542)  SERHVLLYGTNPLSRLKLDNGKLAGLLAQRMFEMLPCGGVGVDTDTIWNE
      SC_HDA1   (119)  LKIPVRAATSEEILEVHTKEHLEFIESTEKMSRE-ELLKETEKGDSVYFN

                         801                                                850
 AQUIFEX_HDAL    (92)  ENPVSYAMFTGSSLATGSTVQAIEEFLKGNVAFNPAGGMHHAFKSRANGF
    BMY_HDAL1   (106)  --------------------------------------------------
    BMY_HDAL2    (10)  LHSSGAARMAVGCVIELASKVASGELKNGFAVVRPPG--HHAEESTAMGF
    BMY_HDAL3     (1)  --------------------------------------------------
         HDA4   (765)  VHSAGAARLAVGCVVELVFKVATGELKNGFAVVRPPG--HHAEESTPMGF
         HDA5   (795)  MHSSSAVRMAVGCLLELAFKVAAGELKNGFAIIRPPG--HHAEESTAMGF
         HDA7   (592)  LHSSNAARWAAGSVTDLAFKVASRELKNGFAVVRPPG--HHADHSTAMGF
      SC_HDA1   (168)  NDSYASARLPCGGAIEACKAVVEGRVKNSLAVVRPPG--HHAEPQAAGGF

                         851                                                900
 AQUIFEX_HDAL   (142)  GYINNPAVGIEYLRKK----GFKRILYIDLDAHHCDGVQEAFYDTDQVFV
    BMY_HDAL1   (106)  --------------------------------------------------
    BMY_HDAL2    (58)  CFFNSVAITAKYLRDQ---LNISKILIVDLDVHHGNGTQQAFYADPSILY
    BMY_HDAL3     (1)  --------------------------------------------------
         HDA4   (813)  CYFNSVAVAAKLLQQR---LSVSKILIVDWDVHHGNGTQQAFYSDPSVLY
         HDA5   (843)  CFFNSVAITAKLLQQK---LNVGKVLIVDWDIHHGNGTQQAFYNDPSVLY
         HDA7   (640)  CFFNSVAIACRQLQQQSKASKASKILIVDWDVHHGNGTQQTFYQDPSVLY
      SC_HDA1   (216)  CLFSNVAVAAKNILKN-YPESVRRIMILDWDIHHGNGTQKSFYQDDQVLY

                         901                                                950
 AQUIFEX_HDAL   (188)  LSLHQ-SPEYAFPFE-KGFLEEIGEGKGKGYNLNIPLPKG----LNDNEF
    BMY_HDAL1   (106)  --------------------------------------------------
    BMY_HDAL2   (105)  ISLHRYDEGNFFPG--SGAPNEVGTGLGEGYNINIAWTGGLDPPMGDVEY
    BMY_HDAL3     (1)  --------------------------------------------------
         HDA4   (860)  MSLHRYDDGNFFPG--SGAPDEVGTGPGVGFNVNMAFTGGLDPPMGDAEY
         HDA5   (890)  ISLHRYDNGNFFPG--SGAPEEVGGGPGVGYNVNVAWTGGVDPPIGDVEY
         HDA7   (690)  ISLHRHDDGNFFPG--SGAVDEVGAGSGEGFNVNVAWAGGLDPPMGDPEY
      SC_HDA1   (265)  VSLHRFEMGKYYPGTIQGQYDQTGEGKGEGFNCNITWPVG---GVGDAEY
```

FIG. 2A

```
                          951                                              1000
AQUIFEX_HDAL     (232)  LFALEKSLEIVKEVFEPEVYLLQLGTDP--LLEDYLSKFNLSNVAFLKAF
   BMY_HDAL1     (106)  --------------------------------------------------
   BMY_HDAL2     (153)  LEAFRLVLLSL---------------------------------------
   BMY_HDAL3       (1)  ----RTIVKRVAKEFDPDMVLVSAGFDALEGHTPPLGGYKVTAKCFGHLT
        HDA4     (908)  LAAFRTVVMPIASEFAPDVVLVSSGFDAVEGHPTPLGGYNLSARCFGYLT
        HDA5     (938)  LTAFRTVVMPIAHEFSPDVVLVSAGFDAVEGHLSPLGGYSVTARCFGHLT
        HDA7     (738)  LAAFRIVVMPIAREFSPDLVLVSAGFDAAEGHPAPLGGYHVSAKCFGYMT
      SC_HDA1     (312)  MWAFEQVVMPMGREFKPDLVIISSGFDAADG--DTIGQCHVTPSCYGHMT

                          1001                                             1050
AQUIFEX_HDAL     (280)  NIVREVFGEGVYLG-GGGYHPYALARAWTLIWCELSGR---EVPEKLNNK
   BMY_HDAL1     (106)  --------------------------------------------------
   BMY_HDAL2     (164)  --------------------------------------------------
   BMY_HDAL3      (47)  KQLMTLADGRVVLALEGGHDLTAICDASEACVNALLGNELEPLAEDILHQ
        HDA4     (958)  KQLMGLAGGRIVLALEGGHDLTAICDASEACVSALLGNELDPLPEKVLQQ
        HDA5     (988)  RQLMTLAGGRVVLALEGGHDLTAICDASEACVSALLSVELQPLDEAVLQQ
        HDA7     (788)  QQLMNLAGGAVVLALEGGHDLTAICDASEACVAALLGNRVDPLSEEGWKQ
      SC_HDA1     (360)  HMLKSLARGNLCVVLEGGYNLDAIARSALSVAKVLIGEPPDELPDPLSDP

                          1051                                             1100
AQUIFEX_HDAL     (326)  AKELLKSIDFEEFDDEVDRSYMLETLKDPWRGGEVRKEVKDTLEKAKASS
   BMY_HDAL1     (106)  --------------------------------------------------
   BMY_HDAL2     (164)  --------------------------------------------------
   BMY_HDAL3      (97)  SPNMNAVISLQKIIEIQSKYWKSVRMVAVPRGCALAGAQL--QEETETVS
        HDA4    (1008)  RPNANAVRSMEKVMEIHSKYWRCLQRTTSTAGRSLIEAQTCENEEAETVT
        HDA5    (1038)  KPNINAVATLEKVIEIQSKHWSCVQKFAAGLGRSLREAQAGETEEAETVS
        HDA7     (838)  KPQPQCHPLSGGRDPGAQ--------------------------------
      SC_HDA1     (410)  KPE--VIEMIDKVIRLQSKYWNCFRRRHANSGCNFNEPINDSIISKNFPL

                          1101                                             1150
AQUIFEX_HDAL     (376)  --------------------------------------------------
   BMY_HDAL1     (106)  --------------------------------------------------
   BMY_HDAL2     (164)  --------------------------------------------------
   BMY_HDAL3     (145)  ALASLTVDVEQPFAQEDSRTAG----EPMEEEPAL---------------
        HDA4    (1058)  AMASLSVGVKPAEKRPDEEPME--------EEPPL---------------
        HDA5    (1088)  AMALLSVGAEQAQAAAAREHSPRPAEEPMEQEPAL---------------
        HDA7     (856)  --------------------------------------------------
      SC_HDA1     (458)  QKAIRQQQQHYLSDEFNFVTLPLVSMDLPDNTVLCTPNISESNTIIIVVH
```

FIG. 2B

```
Genewise results from HDA5_HUMAN_run2 applied to AC002088
Hit 1: bits = 149
      BAC start:56543
      BAC end:74703
      Protein start:684
      Protein end:788

>Results for GCGPROT:HDA5_HUMAN vs AC002088 (forward) [0]

genewisedb output
Score 149.09 bits over entire alignment.
This will be different from per-alignment scores. See manual for details
For computer parsable output, try genewisedb -help or read the manual
Scores as bits over a synchronous coding model

                     Alignment 1 Score 148.82 (Bits)


HDA5        684 G  V  V  Y  D  T  F  M  L  K  H  Q  C  M  C  G  N  T  H  V
                G  +     Y  D     +  M  L  K  H  Q  C  +  C  G  N  +
                G  I  A  Y  D  P  L  M  L  K  H  Q  C  V  C  G  N  S  T  T
AC002088  56543 ggaattgcctatgaccccttgatgctgaaacaccagtgcgtttgtggcaattccaccacc

                H  P  E  H  A  G  R  I  Q  S  I  W  S  R  L  Q  E  T  G
                H  P  E  H  A  G  R  I  Q  S  I  W  S  R  L  Q  E  T  G
                H  P  E  H  A  G  R  I  Q  S  I  W  S  R  L  Q  E  T  G
                caccctgagcatgctggacgaatacagagtatctggtcacgactgcaagaaactggg

HDA5        723 L  L  S  K  C  E                          R  I  R  G  R  K
                L  L  +  K  C  E                          R  I  +  G  R  K
                L  L  N  K  C  E                          R  I  Q  G  R  K
AC002088  56660 ctgctaaataaatgtgagGTAATCC  Intron 1   CAGcgaattcaaggtcgaaaa
                                     <0-----[56678:69695]-0>

                A  T  L  D
                A  +  L  +
                A  S  L  E
                gccagcctggag

HDA5        739 E  I  Q  T  V  H  S  E  Y  H  T  L  L  Y  G  T  S  P  L  N
                E  I  Q     V  H  S  E  +  H  +  L  L  Y  G  T  +  P  L  +
                E  I  Q  L  V  H  S  E  H  H  S  L  L  Y  G  T  N  P  L  D
AC002088  69726 gaaatacagcttgttcattctgaacatcactcactgttgtatggcaccaacccccctggac

                R  Q  K  L  D  S  K  K  L  L
                   Q  K  L  D     +     L  L
                G  Q  K  L  D  P  R  I  L  L
                ggacagaagctggaccccaggatactccta

HDA5   769                          P  I  S  Q  K  M  Y  A  V  L  P
                                          S  Q  K  +  +  +     L  P
                     G:G[ggt]       D  D  S  Q  K  F  F  S  S  L  P
AC002088  69816 GGTCTGTA  Intron 2   TAGGTgatgactctcaaaagttttttcctcattacct
                   <1-----[69817:74644]-1>
```

FIG. 3A

```
                         C  G  G  I  G  V  D  S
                         C  G  G  +  G  V     +
                         C  G  G  L  G  V  S  T
                         tgtggtggacttggggtaagtaca

HDA5         783 G  I  G  V  D  S
                 G  +  G  V     +
                 G  L  G  V  S  T
AC002088 74686 ggacttggggtaagtaca
```

FIG. 3B

```
MOTIFS FROM: BMY_HDAL1.AA.FASTA

 MISMATCHES: 0


  BMY_HDAL1.AA.FASTA  CHECK: 4620  LENGTH: 105   !

AMIDATION               XG(R,K)(R,K)
                         XG(R)(K)
           48: KCERI       QGRK      ASLEE

 (ABSTRACT FILE: 0009.PDOC)

ASN_GLYCOSYLATION       N~(P)(S,T)~(P)
                          N~P(T)~P
           17: QCVCG        NSTT      HPEHA

 (ABSTRACT FILE: 0001.PDOC)

CAMP_PHOSPHO_SITE       (R,K)2X(S,T)
                        (R,K){2}X(S)
           50: ERIQG       RKAS      LEEIQ

 (ABSTRACT FILE: 0004.PDOC)

CK2_PHOSPHO_SITE        (S,T)X2(D,E)
                         (T)X{2}(E)
           20: CGNST       THPE      HAGRI

                         (S)X{2}(E)
           53: QGRKA       SLEE      IQLVH

 (ABSTRACT FILE: 0006.PDOC)

MYRISTYL                G~(E,D,R,K,H,P,F,Y,W)X2(S,T,A,G,C,N)~(P)
                             G~(E,D,R,K,H,P,F,Y,W)X{2}(T)~P
           16: HQCVC                     GNSTTH                   PEHAG

                             G~(E,D,R,K,H,P,F,Y,W)X{2}(S)~P
          100: SLPCG                     GLGVST

 (ABSTRACT FILE: 0008.PDOC)

PKC_PHOSPHO_SITE        (S,T)X(R,K)
                          (S)X(K)
           89: LLGDD       SQK      FFSSL

 (ABSTRACT FILE: 0005.PDOC)
```

FIG. 4

```
     ValAspSerAspThrIleTrpAsnGluLeuHisSerSerGlyAlaAlaArgMetAlaVal
  1  GTGGACAGTGACACCATTTGGAATGAGCTACACTCGTCCGGTGCTGCACGCATGGCTGTT
     CACCTGTCACTGTGGTAAACCTTACTCGATGTGAGCAGGCCACGACGTGCGTACCGACAA

     GlyCysValIleGluLeuAlaSerLysValAlaSerGlyGluLeuLysAsnGlyPheAla
 61  GGCTGTGTCATCGAGCTGGCTTCCAAAGTGGCCTCAGGAGAGCTGAAGAATGGGTTTGCT
     CCGACACAGTAGCTCGACCGAAGGTTTCACCGGAGTCCTCTCGACTTCTTACCCAAACGA

     ValValArgProProGlyHisHisAlaGluGluSerThrAlaMetGlyPheCysPhePhe
121  GTTGTGAGGCCCCCTGGCCATCACGCTGAAGAATCCACAGCCATGGGGTTCTGCTTTTTT
     CAACACTCCGGGGGACCGGTAGTGCGACTTCTTAGGTGTCGGTACCCCAAGACGAAAAAA

     AsnSerValAlaIleThrAlaLysTyrLeuArgAspGlnLeuAsnIleSerLysIleLeu
181  AATTCAGTTGCAATTACCGCCAAATACTTGAGAGACCAACTAAATATAAGCAAGATATTG
     TTAAGTCAACGTTAATGGCGGTTTATGAACTCTCTGGTTGATTTATATTCGTTCTATAAC

     IleValAspLeuAspValHisHisGlyAsnGlyThrGlnGlnAlaPheTyrAlaAspPro
241  ATTGTAGATCTGGATGTTCACCATGGAAACGGTACCCAGCAGGCCTTTTATGCTGACCCC
     TAACATCTAGACCTACAAGTGGTACCTTTGCCATGGGTCGTCCGGAAAATACGACTGGGG

     SerIleLeuTyrIleSerLeuHisArgTyrAspGluGlyAsnPhePheProGlySerGly
301  AGCATCCTGTACATTTCACTCCATCGCTATGATGAAGGGAACTTTTTCCCTGGCAGTGGA
     TCGTAGGACATGTAAAGTGAGGTAGCGATACTACTTCCCTTGAAAAAGGGACCGTCACCT

     AlaProAsnGluValGlyThrGlyLeuGlyGluGlyTyrAsnIleAsnIleAlaTrpThr
361  GCCCCAAATGAGGTTGGAACAGGCCTTGGAGAAGGGTACAATATAAATATTGCCTGGACA
     CGGGGTTTACTCCAACCTTGTCCGGAACCTCTTCCCATGTTATATTTATAACGGACCTGT

     GlyGlyLeuAspProProMetGlyAspValGluTyrLeuGluAlaPheArgLeuValLeu
421  GGTGGCCTTGATCCTCCCATGGGAGATGTTGAGTACCTTGAAGCATTCAGGTTGGTACTT
     CCACCGGAACTAGGAGGGTACCCTCTACAACTCATGGAACTTCGTAAGTCCAACCATGAA

     LeuSerLeu
481  CTTTCTCTC
     GAAAGAGAG
```

FIG. 5

```
GENEWISE RESULTS FROM HDA5_HUMAN_RUN3 APPLIED TO AC002410
HIT 1: BITS = 262
       BAC START:15451
       BAC END:58122
       PROTEIN START:786
       PROTEIN END:948

>RESULTS FOR GCGPROT:HDA5_HUMAN VS AC002410 (FORWARD) [0]

GENEWISEDB OUTPUT
SCORE 262.30 BITS OVER ENTIRE ALIGNMENT.
THIS WILL BE DIFFERENT FROM PER-ALIGNMENT SCORES. SEE MANUAL FOR DETAILS
FOR COMPUTER PARSABLE OUTPUT, TRY GENEWISEDB -HELP OR READ THE MANUAL
SCORES AS BITS OVER A SYNCHRONOUS CODING MODEL

                       ALIGNMENT 1 SCORE 261.25 (BITS)


HDA5       786 V  D  S  D  T  V  W  N  E  M  H  S  S  S  A  V  R  M  A  V  G  C  L
               V  D  S  D  T  +  W  N  E  +  H  S  S     A     R  M  A  V  G  C  +
               V  D  S  D  T  I  W  N  E  L  H  S  S  G  A  A  R  M  A  V  G  C  V
AC002410 15451 GTGGACAGTGACACCATTTGGAATGAGCTACACTCGTCCGGTGCTGCACGCATGGCTGTTGGCTGTGTC

               L  E  L  A  F  K  V  A  A  G  E  L  K
               +  E  L  A     K  V  A  +  G  E  L  K
               I  E  L  A  S  K  V  A  S  G  E  L  K
               ATCGAGCTGGCTTCCAAAGTGGCCTCAGGAGAGCTGAAG

HDA5  822                               N  G  F  A  I  I  R  P  P  G  H  H  A  E  E  S
                                        N  G  F  A  +  +  R  P  P  G  H  H  A  E  E  S
                                        N  G  F  A  V  V  R  P  P  G  H  H  A  E  E  S
AC002410 15559 GTGAGGT   INTRON 1    CAGAATGGGTTTGCTGTTGTGAGGCCCCCTGGCCATCACGCTGAAGAATCC
                      <0-----[15559:51266]-0>

HDA5       838 T  A                             G  F  C  F  F  N  S  V  A  I  T
               T  A                             G  F  C  F  F  N  S  V  A  I  T
               T  A              M:M[ATG]       G  F  C  F  F  N  S  V  A  I  T
AC002410 51315 ACAGCCATGTAAGTA   INTRON 2    CAGGGGGTTCTGCTTTTTTAATTCAGTTGCAATTACC
                           <2-----[51323:51566]-2>

HDA5       852 A  K  L  L  Q  Q  K  L  N  V  G  K  V  L  I  V  D  W
               A  K     L  +     +  L  N  +     K  +  L  I  V  D
               A  K  Y  L  R  D  Q  L  N  I  S  K  I  L  I  V  D  L
AC002410 51601 GCCAAATACTTGAGAGACCAACTAAATATAAGCAAGATATTGATTGTAGATCTGGTATGTA   INTRON 3
                                                                       <0---[51655:57572]

HDA5        870   D  I  H  H  G  N  G  T  Q  Q  A  F  Y  N  D  P  S  V  L  Y  I  S  L
                  D  +  H  H  G  N  G  T  Q  Q  A  F  Y     D  P  S  +  L  Y  I  S  L
                  D  V  H  H  G  N  G  T  Q  Q  A  F  Y  A  D  P  S  I  L  Y  I  S  L
AC002410  57570 TAGGATGTTCACCATGGAAACGGTACCCAGCAGGCCTTTTATGCTGACCCCAGCATCCTGTACATTTCACTC
               -0>
                  H  R  Y  D  N  G  N  F  F  P  G  S  G
                  H  R  Y  D     G  N  F  F  P  G  S  G
                  H  R  Y  D  E  G  N  F  F  P  G  S  G
                  CATCGCTATGATGAAGGGAACTTTTTCCCTGGCAGTGGA

HDA5       906 A  P  E  E                           V  G  G  G  P  G  V  G  Y  N  V  N
               A  P     E                           V  G     G     G     G  Y  N  +  N
               A  P  N  E                           V  G  T  G  L  G  E  G  Y  N  I  N
AC002410 57681 GCCCCAAATGAGGTTCGGT   INTRON 4    CAGGTTGGAACAGGCCTTGGAGAAGGGTACAATATAAAT
                             <0-----[57693:58005]-0>
```

FIG. 6A

```
HDA5        922 V  A  W  T  G  G  V  D  P  P  I  G  D  V  E  Y  L  T  A  F  R  T  V  V
                +  A  W  T  G  G  +  D  P  P  +  G  D  V  E  Y  L     A  F  R     V  +
                I  A  W  T  G  G  L  D  P  P  M  G  D  V  E  Y  L  E  A  F  R  L  V  L
AC002410  58042 ATTGCCTGGACAGGTGGCCTTGATCCTCCCATGGGAGATGTTGAGTACCTTGAAGCATTCAGGTTGGTACTT

                M  P  I
                +     +
                L  S  L
                CTTTCTCTC
```

FIG. 6B

```
PROSITE motifs identified in the partial predicted amino acid sequence of
BMY_HDAL2.
MOTIFS FROM: BMY_HDAL2.AA.FASTA

 MISMATCHES: 0


  BMY_HDAL2.AA.FASTA  CHECK: 2381  LENGTH: 163   !

ASN_GLYCOSYLATION       N~(P)(S,T)~(P)
                           N~P(S)~P
            75: LRDQL        NISK       ILIVD

                           N~P(T)~P
            90: DVHHG        NGTQ       QAFYA

 (ABSTRACT FILE: 0001.PDOC)

MYRISTYL                G~(E,D,R,K,H,P,F,Y,W)X2(S,T,A,G,C,N)~(P)
                             G~(E,D,R,K,H,P,F,Y,W)X{2}(A)~P
            91: VHHGN                   GTQQAF                  YADPS

                             G~(E,D,R,K,H,P,F,Y,W)X{2}(G)~P
           126: APNEV                   GTGLGE                  GYNIN

                             G~(E,D,R,K,H,P,F,Y,W)X{2}(G)~P
           128: NEVGT                   GLGEGY                  NINIA

 (ABSTRACT FILE: 0008.PDOC)

PKC_PHOSPHO_SITE        (S,T)X(R,K)
                         (T)X(K)
            66: NSVAI      TAK      YLRDQ

 (ABSTRACT FILE: 0005.PDOC)
```

FIG. 7

```
GENEWISE RESULTS FROM HDA5_HUMAN_RUN3 APPLIED TO AC004994
HIT 1: BITS = 176
      BAC START:79767
      BAC END:11
      PROTEIN START:942
      PROTEIN END:1055

>RESULTS FOR GCGPROT:HDA5_HUMAN VS AC004994 (REVERSE) [0]

GENEWISEDB OUTPUT
SCORE 176.62 BITS OVER ENTIRE ALIGNMENT.
THIS WILL BE DIFFERENT FROM PER-ALIGNMENT SCORES. SEE MANUAL FOR DETAILS
FOR COMPUTER PARSABLE OUTPUT, TRY GENEWISEDB -HELP OR READ THE MANUAL
SCORES AS BITS OVER A SYNCHRONOUS CODING MODEL

                    ALIGNMENT 1 SCORE 174.85 (BITS)

HDA5_HUMAN  942 R  T  V  V  M  P  I  A  H  E  F  S  P  D  V  V  L  V  S  A  G  F  D  A
                R  T  +  V     P  +  A     E  F     P  D  +  V  L  V  S  A  G  F  D  A
                R  T  I  V  K  P  V  A  K  E  F  D  P  D  M  V  L  V  S  A  G  F  D  A
AC004994 -79767 AGGACCATCGTGAAGCCTGTGGCCAAAGAGTTTGATCCAGACATGGTCTTAGTATCTGCTGGATTTGATGCA
                V  E  G  H  L  S  P  L  G  G  Y  S  V  T  A
                +  E  G  H        P  L  G  G  Y     V  T  A
                L  E  G  H  T  P  P  L  G  G  Y  K  V  T  A
                TTGGAAGGCCACACCCCTCCTCTAGGAGGGTACAAAGTGACGGCA

HDA5_HUMAN  981 R                               F  G  H  L  T  R  Q  L  M  T  L  A
                   +                            F  G  H  L  T  +  Q  L  M  T  L  A
                   K               C:C[TGT]     F  G  H  L  T  K  Q  L  M  T  L  A
AC004994 -79650 AAATGTAAGTA  INTRON 1   TAGGTTTTGGTCATTTGACGAAGCAATTGATGACATTGGCT
                           <1-----[79646:18435]-1>

HDA5_HUMAN  995 G  G  R  V  V  L  A  L  E  G  G  H  D  L  T  A  I  C  D  A  S  E  A  C
              G  R  V  V  L  A  L  E  G  G  H  D  L  T  A  I  C  D  A  S  E  A  C
                D  G  R  V  V  L  A  L  E  G  G  H  D  L  T  A  I  C  D  A  S  E  A  C
AC004994 -18396 GATGGACGTGTGGTGTTGGCTCTAGAAGGAGGACATGATCTCACAGCCATCTGTGATGCATCAGAAGCCTGT
                V  S  A  L  L  S  V  E
                V  +  A  L  L        E
                V  N  A  L  L  G  N  E
                GTAAATGCCCTTCTAGGAAATGAG

HDA5_HUMAN 1027                       L  Q  P  L  D  E  A  V  L  Q  Q  K  P  N  I  N
                                     L  +  P  L     E     +  L     Q     P  N  +  N
                                     L  E  P  L  A  E  D  I  L  H  Q  S  P  N  M  N
AC004994 -18300 GTAAAAA  INTRON 2   CAGCTGGAGCCACTTGCAGAAGATATTCTCCACCAAAGCCCGAATATGAAT
                <0-----[18300:   98]-0>

HDA5_HUMAN 1043 A  V  A  T  L  E  K  V  I  E  I  Q  S
                A  V     +  L  +  K  +  I  E  I  Q  S
                A  V  I  S  L  Q  K  I  I  E  I  Q  S
AC004994    -49 GCTGTTATTTCTTTACAGAAGATCATTGAAATTCAAAGT
```

FIG. 8A

```
GENEWISE RESULTS FROM HDA5_HUMAN_RUN3 APPLIED TO AC004744
HIT 1: BITS = 57
      BAC START:85491
      BAC END:43563
      PROTEIN START:1022
      PROTEIN END:1122

>RESULTS FOR GCGPROT:HDA5_HUMAN VS AC004744 (REVERSE) [0]

GENEWISEDB OUTPUT
SCORE 57.38 BITS OVER ENTIRE ALIGNMENT.
THIS WILL BE DIFFERENT FROM PER-ALIGNMENT SCORES. SEE MANUAL FOR DETAILS
FOR COMPUTER PARSABLE OUTPUT, TRY GENEWISEDB -HELP OR READ THE MANUAL
SCORES AS BITS OVER A SYNCHRONOUS CODING MODEL

                     ALIGNMENT 1 SCORE 55.39 (BITS)

HDA5 1022        L  L  S  V  E  L  Q  P  L  D  E  A  V  L  Q  Q  K  P  N
                 L  L     +  +  L  +  P  L     E     +  L     Q     P  N
                 L  L  F  L  Q  L  E  P  L  A  E  D  I  L  H  Q  S  P  N
AC004744 -85491 CTACTATTCTTGCAGCTGGAGCCACTTGCAGAAGATATTCTCCACCAAAGCCCGAAT

                 I  N  A  V  A  T  L  E  K  V  I  E  I  Q
                 +  N  A  V     +  L  +  K  +  I  E  I  Q
                 M  N  A  V  I  S  L  Q  K  I  I  E  I  Q
                ATGAATGCTGTTATTTCTTTACAGAAGATCATTGAAATTCAA

HDA5 1055                                K  H  W  S  C  V  Q  K  F  A  A  G  L
                                         K  +  W        V  +        A
                         S:S[AGC]        K  Y  W  K  S  V  R  M  V  A  V  P  R
AC004744 85392 AGTATGTC  INTRON 1   TAGGCAAGTATTGGAAGTCAGTAAGGATGGTGGCTGTGCCAAGG
                <1-----[85391:63817]-1>

HDA5 1069       G  R  S  L  R  E  A  Q  A  GET  E  E  A  E  T  V  S  A  M
                G     +  L        A  Q          E  E     E  T  V  S  A  +
                G  C  A  L  A  G  A  Q  L  --Q  E  E  T  E  T  V  S  A  L
AC004744 -63775 GGCTGTGCTCTGGCTGGTGCTCAGTTG  CAAGAGGAGACAGAGACCGTTTCTGCCCTG

                A  L  L  S  V  G  A  E  Q  A  Q  A  AAARE  H
                A     L  +  V        E  Q        A
                A  S  L  T  V  D  V  E  Q  P  F  A  ----Q  E
                GCCTCCCTAACAGTGGATGTGGAACAGCCCTTTGCT     CAGGAA

HDA5 1108       S  P                                P  A  E  E  P  M  E  Q  E  P  A  L
                                                       A     E  P  M  E  +  E  P  A  L
                D  S              R:R[AGA]          T  A  G  E  P  M  E  E  E  P  A  L
AC004744 -63676 GACAGCAGGTATGAA   INTRON 2    CAGAACTGCTGGTGAGCCTATGGAAGAGGAGCCAGCCTTG
                          <2-----[63668:43600]-2>
```

FIG. 8B

```
                       1                                                50
»  AC004744     (1)
»  AC004994     (1)   aggaccatcgtgaagcctgtggccaaagagtttgatccagacatggtct
   BMY_HDAL3    (1)   aggaccatcgtgaagcctgtggccaaagagtttgatccagacatggtct

                     51                                                100
»  AC004744     (1)
»  AC004994    (50)  tagtatctgctggatttgatgcattggaaggccacacccctcctctagga
   BMY_HDAL3   (50)  tagtatctgctggatttgatgcattggaaggccacacccctcctctagga

                     101                                               150
»  AC004744     (1)
»  AC004994   (100)  gggtacaaagtgacggcaaaatgttttggtcatttgacgaagcaattgat
   BMY_HDAL3  (100)  gggtacaaagtgacggcaaaatgttttggtcatttgacgaagcaattgat

                     151                                               200
»  AC004744     (1)
»  AC004994   (150)  gacattggctgatggacgtgtggtgttggctctagaaggaggacatgatc
   BMY_HDAL3  (150)  gacattggctgatggacgtgtggtgttggctctagaaggaggacatgatc

                     201                                               250
»  AC004744     (1)
»  AC004994   (200)  tcacagccatctgtgatgcatcagaagcctgtgtaaatgcccttctagga
   BMY_HDAL3  (200)  tcacagccatctgtgatgcatcagaagcctgtgtaaatgcccttctagga

                     251                                               300
»  AC004744     (1)      agctggagccacttgcagaagatattctccaccaaagcccgaatat
»  AC004994   (250)  aatgagctggagccacttgcagaagatattctccaccaaagcccgaatat
   BMY_HDAL3  (250)  aatgagctggagccacttgcagaagatattctccaccaaagcccgaatat

                     301                                               350
»  AC004744    (50)  gaatgctgttatttctttacagaagatcattgaaattcaaagcaagtatt
»  AC004994   (300)  gaatgctgttatttctttacagaagatcattgaaattcaaa
   BMY_HDAL3  (300)  gaatgctgttatttctttacagaagatcattgaaattcaaagcaagtatt

                     351                                               400
»  AC004744   (100)  ggaagtcagtaaggatggtggctgtgccaaggggctgtgctctggctggt
»  AC004994   (340)
   BMY_HDAL3  (350)  ggaagtcagtaaggatggtggctgtgccaaggggctgtgctctggctggt

                     401                                               450
»  AC004744   (150)  gctcagttgcaagaggagacagagaccgtttctgccctggcctccctaac
»  AC004994   (340)
   BMY_HDAL3  (400)  gctcagttgcaagaggagacagagaccgtttctgccctggcctccctaac

                     451                                               500
»  AC004744   (200)  agtggatgtggaacagccctttgctcaggaagacagcagaactgctggtg
»  AC004994   (340)
   BMY_HDAL3  (450)  agtggatgtggaacagccctttgctcaggaagacagcagaactgctggtg

                     501                     525
»  AC004744   (250)  agcctatggaagaggagccagcctt
»  AC004994   (340)
   BMY_HDAL3  (500)  agcctatggaagaggagccagcctt
```

FIG. 9

```
    ArgThrIleValLysProValAlaLysGluPheAspProAspMetValLeuValSerAla
1   AGGACCATCGTGAAGCCTGTGGCCAAAGAGTTTGATCCAGACATGGTCTTAGTATCTGCT
    TCCTGGTAGCACTTCGGACACCGGTTTCTCAAACTAGGTCTGTACCAGAATCATAGACGA

    GlyPheAspAlaLeuGluGlyHisThrProProLeuGlyGlyTyrLysValThrAlaLys
61  GGATTTGATGCATTGGAAGGCCACACCCCTCCTCTAGGAGGGTACAAAGTGACGGCAAAA
    CCTAAACTACGTAACCTTCCGGTGTGGGGAGGAGATCCTCCCATGTTTCACTGCCGTTTT

    CysPheGlyHisLeuThrLysGlnLeuMetThrLeuAlaAspGlyArgValValLeuAla
121 TGTTTTGGTCATTTGACGAAGCAATTGATGACATTGGCTGATGGACGTGTGGTGTTGGCT
    ACAAAACCAGTAAACTGCTTCGTTAACTACTGTAACCGACTACCTGCACACCACAACCGA

    LeuGluGlyGlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaCysValAsnAla
181 CTAGAAGGAGGACATGATCTCACAGCCATCTGTGATGCATCAGAAGCCTGTGTAAATGCC
    GATCTTCCTCCTGTACTAGAGTGTCGGTAGACACTACGTAGTCTTCGGACACATTTACGG

    LeuLeuGlyAsnGluLeuGluProLeuAlaGluAspIleLeuHisGlnSerProAsnMet
241 CTTCTAGGAAATGAGCTGGAGCCACTTGCAGAAGATATTCTCCACCAAAGCCCGAATATG
    GAAGATCCTTTACTCGACCTCGGTGAACGTCTTCTATAAGAGGTGGTTTCGGGCTTATAC

    AsnAlaValIleSerLeuGlnLysIleIleGluIleGlnSerLysTyrTrpLysSerVal
301 AATGCTGTTATTTCTTTACAGAAGATCATTGAAATTCAAAGCAAGTATTGGAAGTCAGTA
    TTACGACAATAAAGAAATGTCTTCTAGTAACTTTAAGTTTCGTTCATAACCTTCAGTCAT

    ArgMetValAlaValProArgGlyCysAlaLeuAlaGlyAlaGlnLeuGlnGluGluThr
361 AGGATGGTGGCTGTGCCAAGGGGCTGTGCTCTGGCTGGTGCTCAGTTGCAAGAGGAGACA
    TCCTACCACCGACACGGTTCCCCGACACGAGACCGACCACGAGTCAACGTTCTCCTCTGT

    GluThrValSerAlaLeuAlaSerLeuThrValAspValGluGlnProPheAlaGlnGlu
421 GAGACCGTTTCTGCCCTGGCCTCCCTAACAGTGGATGTGGAACAGCCCTTTGCTCAGGAA
    CTCTGGCAAAGACGGGACCGGAGGGATTGTCACCTACACCTTGTCGGGAAACGAGTCCTT

    AspSerArgThrAlaGlyGluProMetGluGluGluProAlaLeu
481 GACAGCAGAACTGCTGGTGAGCCTATGGAAGAGGAGCCAGCCTTG
    CTGTCGTCTTGACGACCACTCGGATACCTTCTCCTCGGTCGGAAC
```

FIG. 10

```
PROSITE MOTIFS FROM: BMY_HDAL3.AA.FASTA

 MISMATCHES:0

  BMY_HDAL3.AA.FASTA  CHECK: 3930  LENGTH: 175   !

CK2_PHOSPHO_SITE       (S,T)X2(D,E)
                        (T)X{2}(D)
             51: TKQLM     TLAD      GRVVL

                        (T)X{2}(E)
            164: QEDSR     TAGE      PMEEE

 (ABSTRACT FILE: 0006.PDOC)

MYRISTYL               G~(E,D,R,K,H,P,F,Y,W)X2(S,T,A,G,C,N)~(P)
                            G~(E,D,R,K,H,P,F,Y,W)X{2}(A)~P
            128: VAVPR                  GCALAG                     AQLQE

 (ABSTRACT FILE: 0008.PDOC)

PKC_PHOSPHO_SITE       (S,T)X(R,K)
                         (T)X(K)
             38: GGYKV     TAK      CFGHL

                         (S)X(R)
            119: SKYWK     SVR      MVAVP

 (ABSTRACT FILE: 0005.PDOC)
```

FIG. 11

Multiple sequence alignment of BMY_HDAL3, AAC78618 and AAD15364

```
                   1                                                50
AAC78618     (1)   -TIVKPVAKEFDPDMVLVSAGFDALEGHTPPLGGYKVTAKCFGHLTKQLM
AAD15364     (1)   --------------------------------------------------
BMY_HDAL3    (1)   RTIVKPVAKEFDPDMVLVSAGFDALEGHTPPLGGYKVTAKCFGHLTKQLM

                   51                                              100
AAC78618    (50)   TLADGRVVLALEGGHDLTAICDASEACVNALLGNELEPLAEDILHQSPNM
AAD15364     (1)   -----------------------------------LEPLAEDILHQSPNM
BMY_HDAL3   (51)   TLADGRVVLALEGGHDLTAICDASEACVNALLGNELEPLAEDILHQSPNM

                   101                                             150
AAC78618   (100)   NAVISLQKIIEIQ-------------------------------------
AAD15364    (16)   NAVISLQKIIEIQKLLVSLWKRSQPCEVPSPPLIFPVCDIIVYPPTPVPS
BMY_HDAL3  (101)   NAVISLQKIIEIQSKYWKSVRMVAVPRGCALAGAQLQEETETVSALASLT

                   151                     175
AAC78618   (113)   -------------------------
AAD15364    (66)   DMSCLLPGWHRFNGT----------
BMY_HDAL3  (151)   VDVEQPFAQEDSRTAGEPMEEEPAL
```

FIG. 12

BLASTN alignment of AA287983 and BMY_HDAL3

```
SCORE =  224 BITS (113), EXPECT = 4E-57
 IDENTITIES = 120/121 (99%), GAPS = 1/121 (0%)
 STRAND = PLUS / MINUS

BMY_HDAL3: 405 ATTTTGCCGTCACTTTGTACCCTCCTAGAGGAGGGGTGTGGCCTTCCAATGCATCAAATC
464
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AA287983:  207 ATTTTGCCGTCACTTTGTACCCTCCTAGAGGAGGGGTGTGGCCTTCCAATGCATCAAATC
148

BMY_HDAL3: 465 CAGCAGATACTAAGACCATGTCTGGATCAAACTCTTTGGCCACAGGCTTCACGATGGTCC
524
               ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
AA287983:  147 CAGCAGATACTAAGACCATGTCTGGATCAAACTCTTT-GCCACAGGCTTCACGATGGTCC 89

BMY_HDAL3: 525 T 525
               |
AA287983:  88  T 88
```

FIG. 13

Aquifex ACUC Protein

```
  1 MKKVKLIGTL DYGKYRYPKN HPLKIPRVSL LLRFKDAMNL IDEKELIKSR
 51 PATKEELLLF HTEDYINTLM EAERCQCVPK GAREKYNIGG YENPVSYAMF
101 TGSSLATGST VQAIEEFLKG NVAFNPAGGM HHAFKSRANG FCYINNPAVG
151 IEYLRKKGFK RILYIDLDAH HCDGVQEAFY DTDQVFVLSL HQSPEYAFPF
201 EKGFLEEIGE GKGKGYNLNI PLPKGLNDNE FLFALEKSLE IVKEVFEPEV
251 YLLQLGTDPL LEDYLSKFNL SNVAFLKAFN IVREVFGEGV YLGGGGYHPY
301 ALARAWTLIW CELSGREVPE KLNNKAKELL KSIDFEEFDD EVDRSYMLET
351 LKDPWRGGEV RKEVKDTLEK AKASS
```

FIG. 14A

Saccharomyces Cerevisiae Histone Deacetylase 1

```
  1 MDSVMVKKEV LENPDHDLKR KLEENKEEEN SLSTTSKSKR QVIVPVCMPK
 51 IHYSPLKTGL CYDVRMRYHA KIFTSYFEYI DPHPEDPRRI YRIYKILAEN
101 GLINDPTLSG VDDLGDLMLK IPVRAATSEE ILEVHTKEHL EFIESTEKMS
151 REELLKETEK GDSVYFNNDS YASARLPCGG AIEACKAVVE GRVKNSLAVV
201 RPPGHHAEPQ AAGGFCLFSN VAVAAKNILK NYPESVRRIM ILDWDIHHGN
251 GTQKSFYQDD QVLYVSLHRF EMGKYYPGTI QGQYDQTGEG KGEGFNCNIT
301 WPVGGVGDAE YMWAFEQVVM PMGREFKPDL VIISSGFDAA DGDTIGQCHV
351 TPSCYGHMTH MLKSLARGNL CVVLEGGYNL DAIARSALSV AKVLIGEPPD
401 ELPDPLSDPK PEVIEMIDKV IRLQSKYWNC FRRRHANSGC NFNEPINDSI
451 ISKNFPLQKA IRQQQHYLS DEFNFVTLPL VSMDLPDNTV LCTPNISESN
501 TIIIVVHDTS DIWAKRNVIS GTIDLSSSVI IDNSLDFIKW GLDRKYGIID
551 VNIPLTLFEP DNYSGMITSQ EVLIYLWDNY IKYFPSVAKI AFIGIGDSYS
601 GIVHLLGHRD TRAVTKTVIN FLGDKQLKPL VPLVDETLSE WYFKNSLIFS
651 NNSHQCWKEN ESRKPRKKFG RVLRCDTDGL NNIIEERFEE ATDFILDSFE
701 EWSDEE
```

FIG. 14B

Homo Sapiens Histone Deacetylase 4

```
   1 MSSQSHPDGL SGRDQPVELL NPARVNHMPS TVDVATALPL QVAPSAVPMD
  51 LRLDHQFSLP VAEPALREQQ LQQELLALKQ KQQIQRQILI AEFQRQHEQL
 101 SRQHEAQLHE HIKQQQEMLA MKHQQELLEH QRKLERHRQE QELEKQHREQ
 151 KLQQLKNKEK GKESAVASTE VKMKLQEFVL NKKKALAHRN LNHCISSDPR
 201 YWYGKTQHSS LDQSSPPQSG VSTSYNHPVL GMYDAKDDFP LRKTASEPNL
 251 KLRSRLKQKV AERRSSPLLR RKDGPVVTAL KKRPLDVTDS ACSSAPGSGP
 301 SSPNNSSGSV SAENGIAPAV PSIPAETSLA HRLVAREGSA APLPLYTSPS
 351 LPNITLGLPA TGPSAGTAGQ QDTERLTLPA LQQRLSLFPG THLTPYLSTS
 401 PLERDGGAAH SPLLQHMVLL EQPPAQAPLV TGLGALPLHA QSLVGADRVS
 451 PSIHKLRQHR PLGRTQSAPL PQNAQALQHL VIQQQHQQFL EKHKQQFQQQ
 501 QLQMNKIIPK PSEPARQPES HPEETEEELR EHQALLDEPY LDRLPGQKEA
 551 HAQAGVQVKQ EPIESDEEEA EPPREVEPGQ RQPSEQELLF RQQALLLEQQ
 601 RIHQLRNYQA SMEAAGIPVS FGGHRPLSRA QSSPASATFP VSVQEPPTKP
 651 RFTTGLVYDT LMLKHQCTCG SSSSHPEHAG RIQSIWSRLQ ETGLRGKCEC
 701 IRGRKATLEE LQTVHSEAHT LLYGTNPLNR QKLDSKKLLG SLASVFVRLP
 751 CGGVGVDSDT IWNEVHSAGA ARLAVGCVVE LVFKVATGEL KNGFAVVRPP
 801 GHHAEESTPM GFCYFNSVAV AAKLLQQRLS VSKILIVDWD VHHGNGTQQA
 851 FYSDPSVLYM SLHRYDDGNF FPGSGAPDEV GTGPGVGFNV NMAFTGGLDP
 901 PMGDAEYLAA FRTVVMPIAS EFAPDVVLVS SGFDAVEGHP TPLGGYNLSA
 951 RCFGYLTKQL MGLAGGRIVL ALEGGHDLTA ICDASEACVS ALLGNELDPL
1001 PEKVLQQRPN ANAVRSMEKV MEIHSKYWRC LQRTTSTAGR SLIEAQTCEN
1051 EEAETVTAMA SLSVGVKPAE KRPDEEPMEE EPPL
```

FIG. 14C

Homo Sapiens Histone Deacetylase 5

```
   1 MNSPNESDGM SGREPSLEIL PRTSLHSIPV TVEVKPVLPR AMPSSMGGGG
  51 GGSPSPVELR GALVGSVDPT LREQQLQQEL LALKQQQQLQ KQLLFAEFQK
 101 QHDHLTRQHE VQLQKHLKQQ QEMLAAKQQQ EMLAAKRQQE LEQQRQREQQ
 151 RQEELEKQRL EQQLLILRNK EKSKESAIAS TEVKLRLQEF LLSKSKEPTP
 201 GGLNHSLPQH PKCWGAHHAS LDQSSPPQSG PPGTPPSYKL PLPGPYDSRD
 251 DFPLRKTASE PNLKVRSRLK QKVAERRSSP LLRRKDGTVI STFKKRAVEI
 301 TGAGPGASSV CNSAPGSGPS SPNSSHSTIA ENGFTGSVPN IPTEMLPQHR
 351 ALPLDSSPNQ FSLYTSPSLP NISLGLQATV TVTNSHLTAS PKLSTQQEAE
 401 RQALQSLRQG GTLTGKFMST SSIPGCLLGV ALEGDGSPHG HASLLQHVLL
 451 LEQARQQSTL IAVPLHGQSP LVTGERVATS MRTVGKLPRH RPLSRTQSSP
 501 LPQSPQALQQ LVMQQQHQQF LEKQKQQQLQ LGKILTKTGE LPRQPTTHPE
 551 ETEEELTEQQ EVLLGEGALT MPREGSTESE STQEDLEEED EEEDGEEEED
 601 CIQVKDEEGE SGAEEGPDLE EPGAGYKKLF SDAQPLQPLQ VYQAPLSLAT
 651 VPHQALGRTQ SSPAAPGGMK SPPDQPVKHL FTTGVVYDTF MLKHQCMCGN
 701 THVHPEHAGR IQSIWSRLQE TGLLSKCERI RGRKATLDEI QTVHSEYHTL
 751 LYGTSPLNRQ KLDSKKLLGP ISQKMYAVLP CGGIGVDSDT VWNEMHSSSA
 801 VRMAVGCLLE LAFKVAAGEL KNGFAIIRPP GHHAEESTAM GFCFFNSVAI
 851 TAKLLQQKLN VGKVLIVDWD IHHGNGTQQA FYNDPSVLYI SLHRYDNGNF
 901 FPGSGAPEEV GGGPGVGYNV NVAWTGGVDP PIGDVEYLTA FRTVVMPIAH
 951 EFSPDVVLVS AGFDAVEGHL SPLGGYSVTA RCFGHLTRQL MTLAGGRVVL
1001 ALEGGHDLTA ICDASEACVS ALLSVELQPL DEAVLQQKPN INAVATLEKV
1051 IEIQSKHWSC VQKFAAGLGR SLREAQAGET EEAETVSAMA LLSVGAEQAQ
1101 AAAAREHSPR PAEEPMEQEP AL
```

FIG. 14D

Homo Sapiens Histone Deacetylase 7

```
  1 MDLRVGQRPP VEPPPEPTLL ALQRPQRLHH HLFLAGLQQQ RSVEPMRLSM
 51 DTPMPELQVG PQEQELRQLL HKDKSKRSAV ASSVVKQKLA EVILKKQQAA
101 LERTVHPNSP GIPYRTLEPL ETEGATRSML SSFLPPVPSL PSDPPEHFPL
151 RKTVSEPNLK LRYKPKKSLE RRKNPLLRKE SAPPSLRRRP AETLGDSSPS
201 SSSTPASGCS SPNDSEHGPN PILGDSDRRT HPTLGPRGPI LGSPHTPLFL
251 PHGLEPEAGG TLPSRLQPIL LLDPSGSHAP LLTVPGLGPL PFHFAQSLMT
301 TERLSGSGLH WPLSRTRSEP LPPSATAPPP PGPMQPRLEQ LKTHVQVIKR
351 SAKPSEKPRL RQIPSAEDLE TDGGGPGQVV DDGLEHRELG HGQPEARGPA
401 PLQQHPQVLL WEQQRLAGRL PRGSTGDTVL LPLAQGGHRP LSRAQSSPAA
451 PASLSAPEPA SQARVLSSSE TPARTLPFTT GLIYDSVMLK HQCSCGDNSR
501 HPEHAGRIQS IWSRLQERGL RSQCECLRGR KASLEELQSV HSERHVLLYG
551 TNPLSRLKLD NGKLAGLLAQ RMFEMLPCGG VGVDTDTIWN ELHSSNAARW
601 AAGSVTDLAF KVASRELKNG FAVVRPPGHH ADHSTAMGFC FFNSVAIACR
651 QLQQQSKASK ASKILIVDWD VHHGNGTQQT FYQDPSVLYI SLHRHDDGNF
701 FPGSGAVDEV GAGSGEGFNV NVAWAGGLDP PMGDPEYLAA FRIVVMPIAR
751 EFSPDLVLVS AGFDAAEGHP APLGGYHVSA KCFGYMTQQL MNLAGGAVVL
801 ALEGGHDLTA ICDASEACVA ALLGNRVDPL SEEGWKQKPQ PQCHPLSGGR
851 DPGAQ
```

FIG. 14E

Human EST AA287983

```
  1 ggccttggagaagggtacaatataaatattgcctggacaggtggcctt
 49 gatcctcccatgggagatgttgagtaccttgaagcattcaggaccatc
 97 gtgaagcctgtggcaaagagtttgatccagacatggtcttagtatctg
145 ctggatttgatgcattggaaggccacacccctcctctaggagggtaca
193 aagtgacggcaaaataaactcctgtgctggaggtacaacagtttggaa
241 gtatacttggggaaagagaaaacacaagatggaaggaagatctctctt
289 ttcacatcgggagcac
```

FIG. 14F

Human predicted protein AAD15364

```
 1 LEPLAEDILH QSPNMNAVIS LQKIIEIQKL LVSLWKRSQP CEVPSPPLIF
51 PVCDIIVYPP TPVPSDMSCL LPGWHRFNGT
```

FIG. 14G

Human predicted protein AAC78618

```
  1 TIVKPVAKEF DPDMVLVSAG FDALEGHTPP LGGYKVTAKC FGHLTKQLMT
 51 LADGRVVLAL EGGHDLTAIC DASEACVNAL LGNELEPLAE DILHQSPNMN
101 AVISLQKIIE IQ
```

FIG. 14H

```
   1 ATGCACAGTATGATCAGCTCAGTGGATGTGAAGTCAGAAGTTCCTGTGGGCCTGGAGCCC   60
   1 M  H  S  M  I  S  S  V  D  V  K  S  E  V  P  V  G  L  E  P    20

  61 ATCTCACCTTTAGACCTAAGGACAGACCTCAGGATGATGATGCCCGTGGTGGACCCTGTT  120
  21 I  S  P  L  D  L  R  T  D  L  R  M  M  M  P  V  V  D  P  V    40

 121 GTCCGTGAGAAGCAATTGCAGCAGGAATTACTTCTTATCCAGCAGCAGCAACAAATCCAG  180
  41 V  R  E  K  Q  L  Q  Q  E  L  L  L  I  Q  Q  Q  Q  Q  I  Q    60

 181 AAGCAGCTTCTGATAGCAGAGTTTCAGAAACAGCATGAGAACTTGACACGGCAGCACCAG  240
  61 K  Q  L  L  I  A  E  F  Q  K  Q  H  E  N  L  T  R  Q  H  Q    80

 241 GCTCAGCTTCAGGAGCATATCAAGTTGCAACAGGAACTTCTAGCCATAAAACAGCAACAA  300
  81 A  Q  L  Q  E  H  I  K  L  Q  Q  E  L  L  A  I  K  Q  Q  Q   100

 301 GAACTCCTAGAAAAGGAGCAGAAACTGGAGCAGCAGAGGCAAGAACAGGAAGTAGAGAGG  360
 101 E  L  L  E  K  E  Q  K  L  E  Q  Q  R  Q  E  Q  E  V  E  R   120

 361 CATCGCAGAGAACAGCAGCTTCCTCCTCTCAGAGGCAAAGATAGAGGACGAGAAAGGGCA  420
 121 H  R  R  E  Q  Q  L  P  P  L  R  G  K  D  R  G  R  E  R  A   140

 421 GTGGCAAGTACAGAAGTAAAGCAGAAGCTTCAAGAGTTCCTACTGAGTAAATCAGCAACG  480
 141 V  A  S  T  E  V  K  Q  K  L  Q  E  F  L  L  S  K  S  A  T   160

 481 AAAGACACTCCAACTAATGGAAAAAATCATTCCGTGAGCCGCCATCCCAAGCTCTGGTAC  540
 161 K  D  T  P  T  N  G  K  N  H  S  V  S  R  H  P  K  L  W  Y   180

 541 ACGGCTGCCCACCACACATCATTGGATCAAAGCTCTCCACCCCTTAGTGGAACATCTCCA  600
 181 T  A  A  H  H  T  S  L  D  Q  S  S  P  P  L  S  G  T  S  P   200

 601 TCCTACAAGTACACATTACCAGGAGCACAAGATGCAAAGGATGATTTCCCCCTTCGAAAA  660
 201 S  Y  K  Y  T  L  P  G  A  Q  D  A  K  D  D  F  P  L  R  K   220

 661 ACTGCCTCTGAGCCCAACTTGAAGGTGCGGTCCAGGTTAAAACAGAAAGTGGCAGAGAGG  720
 221 T  A  S  E  P  N  L  K  V  R  S  R  L  K  Q  K  V  A  E  R   240

 721 AGAAGCAGCCCCTTACTCAGGCGGAAGGATGGAAATGTTGTCACTTCATTCAAGAAGCGA  780
 241 R  S  S  P  L  L  R  R  K  D  G  N  V  V  T  S  F  K  K  R   260

 781 ATGTTTGAGGTGACAGAATCCTCAGTCAGTAGCAGTTCTCCAGGCTCTGGTCCCAGTTCA  840
 261 M  F  E  V  T  E  S  S  V  S  S  S  S  P  G  S  G  P  S  S   280

 841 CCAAACAATGGGCCAACTGGAAGTGTTACTGAAAATGAGACTTCGGTTTTGCCCCCTACC  900
 281 P  N  N  G  P  T  G  S  V  T  E  N  E  T  S  V  L  P  P  T   300

 901 CCTCATGCCGAGCAAATGGTTTCACAGCAACGCATTCTAATTCATGAAGATTCCATGAAC  960
 301 P  H  A  E  Q  M  V  S  Q  Q  R  I  L  I  H  E  D  S  M  N   320

 961 CTGCTAAGTCTTTATACCTCTCCTTCTTTGCCCAACATTACCTTGGGGCTTCCCGCAGTG 1020
 321 L  L  S  L  Y  T  S  P  S  L  P  N  I  T  L  G  L  P  A  V   340

1021 CCATCCCAGCTCAATGCTTCGAATTCACTCAAAGAAAAGCAGAAGTGTGAGACGCAGACG 1080
 341 P  S  Q  L  N  A  S  N  S  L  K  E  K  Q  K  C  E  T  Q  T   360

1081 CTTAGGCAAGGTGTTCCTCTGCCTGGGCAGTATGGAGGCAGCATCCCGGCATCTTCCAGC 1140
 361 L  R  Q  G  V  P  L  P  G  Q  Y  G  G  S  I  P  A  S  S  S   380

1141 CACCCTCATGTTACTTTAGAGGGAAAGCCACCCAACAGCAGCCACCAGGCTCTCCTGCAG 1200
 381 H  P  H  V  T  L  E  G  K  P  P  N  S  S  H  Q  A  L  L  Q   400

1201 CATTTATTATTGAAAGAACAAATGCGACAGCAAAAGCTTCTTGTAGCTGGTGGAGTTCCC 1260
 401 H  L  L  L  K  E  Q  M  R  Q  Q  K  L  L  V  A  G  G  V  P   420

1261 TTACATCCTCAGTCTCCCTTGGCAACAAAAGAGAGAATTTCACCTGGCATTAGAGGTACC 1320
 421 L  H  P  Q  S  P  L  A  T  K  E  R  I  S  P  G  I  R  G  T   440

1321 CACAAATTGCCCCGTCACAGACCCCTGAACCGAACCCAGTCTGCACCTTTGCCTCAGAGC 1380
 441 H  K  L  P  R  H  R  P  L  N  R  T  Q  S  A  P  L  P  Q  S   460

1381 ACGTTGGCTCAGCTGGTCATTCAACAGCAACACCAGCAATTCTTGGAGAAGCAGAAGCAA 1440
 461 T  L  A  Q  L  V  I  Q  Q  Q  H  Q  Q  F  L  E  K  Q  K  Q   480
```

FIG. 15A

```
1441  TACCAGCAGCAGATCCACATGAACAAACTGCTTTCGAAATCTATTGAACAACTGAAGCAA  1500
 481  Y  Q  Q  Q  I  H  M  N  K  L  L  S  K  S  I  E  Q  L  K  Q   500
1501  CCAGGCAGTCACCTTGAGGAAGCAGAGGAAGAGCTTCAGGGGGACCAGGCGATGCAGGAA  1560
 501  P  G  S  H  L  E  E  A  E  E  E  L  Q  G  D  Q  A  M  Q  E   520
1561  GACAGAGCGCCCTCTAGTGGCAACAGCACTAGGAGCGACAGCAGTGCTTGTGTGGATGAC  1620
 521  D  R  A  P  S  S  G  N  S  T  R  S  D  S  S  A  C  V  D  D   540
1621  ACACTGGGACAAGTTGGGGCTGTGAAGGTCAAGGAGGAACCAGTGGACAGTGATGAAGAT  1680
 541  T  L  G  Q  V  G  A  V  K  V  K  E  E  P  V  D  S  D  E  D   560
1681  GCTCAGATCCAGGAAATGGAATCTGGGGAGCAGGCTGCTTTTATGCAACAGCCTTTCCTG  1740
 561  A  Q  I  Q  E  M  E  S  G  E  Q  A  A  F  M  Q  Q  P  F  L   580
1741  GAACCCACGCACACACGTGCGCTCTCTGTGCGCCAAGCTCCGCTGGCTGCGGTTGGCATG  1800
 581  E  P  T  H  T  R  A  L  S  V  R  Q  A  P  L  A  A  V  G  M   600
1801  GATGGATTAGAGAAACACCGTCTCGTCTCCAGGACTCACTCTTCCCCTGCTGCCTCTGTT  1860
 601  D  G  L  E  K  H  R  L  V  S  R  T  H  S  S  P  A  A  S  V   620
1861  TTACCTCACCCGGCAATGGACCGCCCCCTCCAGCCTGGCTCTGCAACTGGAATTGCCTAT  1920
 621  L  P  H  P  A  M  D  R  P  L  Q  P  G  S  A  T  G  I  A  Y   640
1921  GACCCCTTGATGCTGAAACACCAGTGCGTTTGTGGCAATTCCACCACCCACCCTGAGCAT  1980
 641  D  P  L  M  L  K  H  Q  C  V  C  G  N  S  T  T  H  P  E  H   660
1981  GCTGGACGAATACAGAGTATCTGGTCACGACTGCAAGAAACTGGGCTGCTAAATAAATGT  2040
 661  A  G  R  I  Q  S  I  W  S  R  L  Q  E  T  G  L  L  N  K  C   680
2041  GAGCGAATTCAAGGTCGAAAAGCCAGCCTGGAGGAAATACAGCTTGTTCATTCTGAACAT  2100
 681  E  R  I  Q  G  R  K  A  S  L  E  E  I  Q  L  V  H  S  E  H   700
2101  CACTCACTGTTGTATGGCACCAACCCCCTGGACGGACAGAAGCTGGACCCCAGGATACTC  2160
 701  H  S  L  L  Y  G  T  N  P  L  D  G  Q  K  L  D  P  R  I  L   720
2161  CTAGGTGATGACTCTCAAAAGTTTTTTTCCTCATTACCTTGTGGTGGACTTGGGGTGGAC  2220
 721  L  G  D  D  S  Q  K  F  F  S  S  L  P  C  G  G  L  G  V  D   740
2221  AGTGACACCATTTGGAATGAGCTACACTCGTCCGGTGCTGCACGCATGGCTGTTGGCTGT  2280
 741  S  D  T  I  W  N  E  L  H  S  S  G  A  A  R  M  A  V  G  C   760
2281  GTCATCGAGCTGGCTTCCAAAGTGGCCTCAGGAGAGCTGAAGAATGGGTTTGCTGTTGTG  2340
 761  V  I  E  L  A  S  K  V  A  S  G  E  L  K  N  G  F  A  V  V   780
2341  AGGCCCCCTGGCCATCACGCTGAAGAATCCACAGCCATGGGGTTCTGCTTTTTTAATTCA  2400
 781  R  P  P  G  H  H  A  E  E  S  T  A  M  G  F  C  F  F  N  S   800
2401  GTTGCAATTACCGCCAAATACTTGAGAGACCAACTAAATATAAGCAAGATATTGATTGTA  2460
 801  V  A  I  T  A  K  Y  L  R  D  Q  L  N  I  S  K  I  L  I  V   820
2461  GATCTGGATGTTCACCATGGAAACGGTACCCAGCAGGCCTTTTATGCTGACCCCAGCATC  2520
 821  D  L  D  V  H  H  G  N  G  T  Q  Q  A  F  Y  A  D  P  S  I   840
2521  CTGTACATTTCACTCCATCGCTATGATGAAGGGAACTTTTTCCCTGGCAGTGGAGCCCCA  2580
 841  L  Y  I  S  L  H  R  Y  D  E  G  N  F  F  P  G  S  G  A  P   860
2581  AATGAGGTTGGAACAGGCCTTGGAGAAGGGTACAATATAAATATTGCCTGGACAGGTGGC  2640
 861  N  E  V  G  T  G  L  G  E  G  Y  N  I  N  I  A  W  T  G  G   880
2641  CTTGATCCTCCCATGGGAGATGTTGAGTACCTTGAAGCATTCAGGACCATCGTGAAGCCT  2700
 881  L  D  P  P  M  G  D  V  E  Y  L  E  A  F  R  T  I  V  K  P   900
2701  GTGGCCAAAGAGTTTGATCCAGACATGGTCTTAGTATCTGCTGGATTTGATGCATTGGAA  2760
 901  V  A  K  E  F  D  P  D  M  V  L  V  S  A  G  F  D  A  L  E   920
2761  GGCCACACCCCTCCTCTAGGAGGGTACAAAGTGACGGCAAAATGTTTTGGTCATTTGACG  2820
 921  G  H  T  P  P  L  G  G  Y  K  V  T  A  K  C  F  G  H  L  T   940
2821  AAGCAATTGATGACATTGGCTGATGGACGTGTGGTGTTGGCTCTAGAAGGAGGACATGAT  2880
 941  K  Q  L  M  T  L  A  D  G  R  V  V  L  A  L  E  G  G  H  D   960
```

FIG. 15B

```
2881  CTCACAGCCATCTGTGATGCATCAGAAGCCTGTGTAAATGCCCTTCTAGGAAATGAGCTG  2940
 961   L  T  A  I  C  D  A  S  E  A  C  V  N  A  L  L  G  N  E  L   980

2941  GAGCCACTTGCAGAAGATATTCTCCACCAAAGCCCGAATATGAATGCTGTTATTTCTTTA  3000
 981   E  P  L  A  E  D  I  L  H  Q  S  P  N  M  N  A  V  I  S  L  1000

3001  CAGAAGATCATTGAAATTCAAAGCAAGTATTGGAAGTCAGTAAGGATGGTGGCTGTGCCA  3060
1001   Q  K  I  I  E  I  Q  S  K  Y  W  K  S  V  R  M  V  A  V  P  1020

3061  AGGGGCTGTGCTCTGGCTGGTGCTCAGTTGCAAGAGGAGACAGAGACCGTTTCTGCCCTG  3120
1021   R  G  C  A  L  A  G  A  Q  L  Q  E  E  T  E  T  V  S  A  L  1040

3121  GCCTCCCTAACAGTGGATGTGGAACAGCCCTTTGCTCAGGAAGACAGCAGAACTGCTGGT  3180
1041   A  S  L  T  V  D  V  E  Q  P  F  A  Q  E  D  S  R  T  A  G  1060

3181  GAGCCTATGGAAGAGGAGCCAGCCTTGTGAAGTGCCAAGTCCCCCTCTGATATTTCCTGT  3240
1061   E  P  M  E  E  E  P  A  L                                    1069

3241  GTGTGACATCATTGTGTATCCCCCCACCCCAGTACCCTCAGACATGTCTTGTCTGCTGCC  3300
3301  TGGGTGGCACAGATTCAATGGAACATAAACACTGGGCACAAAATTCTGAACAGCAGCTTC  3360
3361  ACTTGTTCTTTGGATGGACTTGAAAGGGCATTAAAGATTCCTTAAACGTAACCGCTGTGA  3420
3421  TTCTAGAGTTACAGTAAACCACGATTGGAAGAAACTGCTTCCAGCATGCTTTTAATATGC  3480
3481  TGGGTGACCCACTCCTAGACACCAAGTTTGAACTAGAAACATTCAGTACAGCACTAGATA  3540
3541  TTGTTAATTTCAGAAGCTATGACAGCCAGTGAAATTTTGGGCAAAACCTGAGACATAGTC  3600
3601  ATTCCTGACATTCTGATCAGCTTTTTTTGGGGTAATTTGTTTTTCAAACAGTCTTAACTT  3660
3661  GTTTACAAGATTTGCTTTTAGCTATGAACGGATCGTAATTCCACCCAGAATGTAATGTTT  3720
3721  CTTGTTTGTTTGTTTTGTTTTGTTAGGGTTTTTTTCTCAACTTTAACACACAGTTCAACT  3780
3781  GTTCCTAGTAAAAGTTCAAGATGGAGGAACTAGCATGAGGCTTTTTTCAGTATCTCGAAG  3840
3841  TCCAAATGCCAAAGGAACCTCACACACTGTTTGTAATGGTGCAATATTTTATATCACTTT  3900
3901  TTTTTAAACATCCCCAACATCTTTGTGTTCTCACACACAGGCAATTTGCAATGTTGCAAT  3960
3961  TGTGTTGGAGAATGAAGTCCCCCCACCTCCCAGCCACACACACATCCTTTGTTCTCATGA  4020
4021  CAGTAGGTCTGAGCAAATGTTCCACCAAGCATTTTCAGTGTCTTTGAAAAGCACGTAACT  4080
4081  TTTCAAAGGTGGTCTTAATTTGCTGCATATCTATCAAGGACTTATTCACTCACCTTTCCT  4140
4141  TTTCTGCCCTCTATCAATTGATTTCTTCTTACCTTTCATCATTCATTCCTTCCTTTAGAA  4200
4201  AAACTGAAGATTACCCATAATCTCCTCTTATTACTTGAGGGCCTTGACTATTTAGTTTAT  4260
4261  TTTGTTTACTTTACAGGTTAACACAGTTGTTTTGTCTGATTGCATTTTATTAACTGTGAA  4320
4321  GCCGTTGAAATGAATATCACTTAAGCAACGTTGCTAAATTTCTATGTGTTTGAAATGTGT  4380
4381  TAATGAAGGCACTGCTTATTTGTAGTCACCTTGAACTGACTTAACCTAGAAGCTGTGCCT  4440
4441  TCTTGTGAAAAAAAAAAAAAAAAAAAA  4467
```

FIG. 15C

```
                1                                                 50
  HDAC9c    (1) ------------------------MHSMISSVDVKSEVPVGLEPISPLDL
AY032737    (1) ------------------------MHSMISSVDVKSEVPVGLEPISPLDL
AY032738    (1) ------------------------MHSMISSVDVKSEVPVGLEPISPLDL
AF132608    (1) MNSPNESDGMSGREPSLEILPRTSLHSIPVTVEVKPVLPRAMPSSMGGGG

                51                                               100
  HDAC9c   (27) RTDLR-------MMMPVVDPVVREKQLQQELLLIQQQQIQKQLLIAEFQK
AY032737   (27) RTDLR-------MMMPVVDPVVREKQLQQELLLIQQQQIQKQLLIAEFQK
AY032738   (27) RTDLR-------MMMPVVDPVVREKQLQQELLLIQQQQIQKQLLIAEFQK
AF132608   (51) GGSPSPVELRGALVGSVDPTLREQQLQQELLALKQQQQEQKQLLFAEFQK

                101                                              150
  HDAC9c   (71) QHENLTRQHQAQLQEHIKLQQELLAIKQQQELLEK--EQKLEQQRQ----
AY032737   (71) QHENLTRQHQAQLQEHIK---ELLAIKQQQELLEK--EQKLEQQRQ----
AY032738   (71) QHENLTRQHQAQLQEHIK---ELLAIKQQQELLEK--EQKLEQQRQ----
AF132608  (101) QHDHLTRQHEVQLQKHLKQQQEMLAAKQQQEMLAAKRQQELEQQRQREQQ

                151                                              200
  HDAC9c  (115) -EQEVERHRREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDT
AY032737  (112) -EQEVERHRREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDT
AY032738  (112) -EQEVERHRREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDT
AF132608  (151) RQEELEKQRLEQQLLILRNKEKSKESAIASTEVKLRLQEFLLSK--SKEP

                201                                              250
  HDAC9c  (164) PTNGKNHSVSRHPKLWYTAAHHTSLDQSSPP---LSGTSPSYKYTLPGAQ
AY032737  (161) PTNGKNHSVSRHPKLWYTAAHHTSLDQSSPP---LSGTSPSYKYTLPGAQ
AY032738  (161) PTNGKNHSVSRHPKLWYTAAHHTSLDQSSPP---LSGTSPSYKYTLPGAQ
AF132608  (199) TPGGLNHSLPQHPKCWG--AHHASLDQSSPPQSGPPGTPPSYKLPLPGPY

                251                                              300
  HDAC9c  (211) DAKDDFPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGNVVTSFKKR
AY032737  (208) DAKDDFPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGNVVTSFKKR
AY032738  (208) DAKDDFPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGNVVTSFKKR
AF132608  (247) DSRDDFPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGTVISTFKKR

                301                                              350
  HDAC9c  (261) MFEVT-----ESSVSSSSPGSGPSSPNNGPTGSVTENETSVLPPTPHAEQ
AY032737  (258) MFEVT-----ESSVSSSSPGSGPSSPNNGPTGSVTENETSVLPPTPHAEQ
AY032738  (258) MFEVT-----ESSVSSSSPGSGPSSPNNGPTGSVTENETSVLPPTPHAEQ
AF132608  (297) AVEITGAGPGASSVCNSAPGSGPSSPNS--SHSTIAENGFTGSVPNIPTE

                351                                              400
  HDAC9c  (306) MVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQLNASNSLK----
AY032737  (303) MVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQLNASNSLK----
AY032738  (303) MVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQLNASNSLK----
AF132608  (345) MLPQHRALPLDSSPNQFSLYTSPSLPNISLGLQATVTVTNSHLTASPKLS
```

FIG. 15D

```
                   401                                              450
  HDAC9c   (352) ---EKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQAL
AY032737   (349) ---EKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQAL
AY032738   (349) ---EKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQAL
AF132608   (395) TQQEAERQALQSLRQGGTLTGKFMSTSSIPGCLLGVALEGDGSPHGHASL

                   451                                              500
  HDAC9c   (399) LQHLLLKEQMRQQKLLVAGGVPLHPQSPLATKERISPGIRGTHKLPRHRP
AY032737   (396) LQHLLLKEQMRQQKLLVAGGVPLHPQSPLATKERISPGIRGTHKLPRHRP
AY032738   (396) LQHLLLKEQMRQQKLLVAGGVPLHPQSPLATKERISPGIRGTHKLPRHRP
AF132608   (445) LQHVLLLEQARQQ--STLIAVPLHGQSPLVTGERVATSMRTVGKLPRHRP

                   501                                              550
  HDAC9c   (449) LNRTQSAPLPQS--TLAQLVIQQQHQQFLEKQKQYQQQIHMNKLLSKSIE
AY032737   (446) LNRTQSAPLPQS--TLAQLVIQQQHQQFLEKQKQYQQQIHMNKLLSKSIE
AY032738   (446) LNRTQSAPLPQS--TLAQLVIQQQHQQFLEKQKQYQQQIHMNKLLSKSIE
AF132608   (493) LSRTQSSPLPQSPQALQQLVMQQQHQQFLEKQK--QQQLQLGKILTKTGE

                   551                                              600
  HDAC9c   (497) QLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDSSACVDDTLGQVG
AY032737   (494) QLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDSSACVDDTLGQVG
AY032738   (494) QLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDSSACVDDTLGQVG
AF132608   (541) LPRQPTTHPEETEEELTEQQEVLLGEGALIMPREGSTESESTQEDLEEED

                   601                                              650
  HDAC9c   (547) AVKVKEEPVDSDEDAQIQEMESGEQAAFMQQP-------FLEPTHTRALS
AY032737   (544) AVKVKEEPVDSDEDAQIQEMESGEQAAFMQQP-------FLEPTHTRALS
AY032738   (544) AVKVKEEPVDSDEDAQIQEMESGEQAAFMQQP-------FLEPTHTRALS
AF132608   (591) EEEDGEEEEDCIQVKDEEGESGAEEGPDLEEPGAGYKKLFSDAQPLQPLQ

                   651                                              700
  HDAC9c   (590) VRQAPLAAVGMDGLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGSATGIA
AY032737   (587) VRQAPLAAVGMDGLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGSATGIA
AY032738   (587) VRQAPLAAVGMDGLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGSATGIA
AF132608   (641) VYQAPLSLATVP----HQALGRTQSSPAAPGGMKSPPDQPVKHLFTTGVV

                   701                                              750
  HDAC9c   (640) YDPLMLKHQCVCGNSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKAS
AY032737   (637) YDPLMLKHQCVCGNSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKAS
AY032738   (637) YDPLMLKHQCVCGNSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKAS
AF132608   (687) YDTFMLKHQCMCGNTHVHPEHAGRIQSIWSRLQETGLLSKCERIRGRKAT

                   751                                              800
  HDAC9c   (690) LEEIQLVHSEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLPCGGLGV
AY032737   (687) LEEIQLVHSEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLPCGGLGV
AY032738   (687) LEEIQLVHSEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLPCGGLGV
AF132608   (737) LDEIQTVHSEYHTLLYGTSPLNRQKLDSKKLLGPISQKMYAVLPCGGIGV

                   801                                              850
  HDAC9c   (740) DSDTIWNELHSSGAARMAVGCVIELASKVASGELKNGFAVVRPPGHHAEE
AY032737   (737) DSDTIWNELHSSGAARMAVGCVIELASKVASGELKNGFAVVRPPGHHAEE
AY032738   (737) DSDTIWNELHSSGAARMAVGCVIELASKVASGELKNGFAVVRPPGHHAEE
AF132608   (787) DSDTVWNEMHSSSAVRMAVGCLLELAFKVAAGELKNGFAIIRPPGHHAEE
```

FIG. 15E

```
                     851                                              900
   HDAC9c    (790)   STAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNGTQQAFYADPS
 AY032737    (787)   STAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNGTQQAFYADPS
 AY032738    (787)   STAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNGTQQAFYADPS
 AF132608    (837)   STAMGFCFFNSVAITAKLLQQKLNVGKVLIVDWDIHHGNGTQQAFYNDPS

                     901                                              950
   HDAC9c    (840)   ILYISLHRYDEGNFFPGSGAPNEVGTGLGEGYNINIAWTGGLDPPMGDVE
 AY032737    (837)   ILYISLHRYDEGNFFPGSGAPNEVGTGLGEGYNINIAWTGGLDPPMGDVE
 AY032738    (837)   ILYISLHRYDEGNFFPGSGAPNEVRFISLEPHFYLYLSGNCIA-------
 AF132608    (887)   VLYISLHRYDNGNFFPGSGAPEEVGGGPGVGYNVNVAWTGGVDPPIGDVE

                     951                                             1000
   HDAC9c    (890)   YLEAFRTIVKPVAKEFDPDMVLVSAGFDALEGHTPPLGGYKVTAKCFGHL
 AY032737    (887)   YLEAFRTIVKPVAKEFDPDMVLVSAGFDALEGHTPPLGGYKVTAKCFGHL
 AY032738    (880)   --------------------------------------------------
 AF132608    (937)   YLTAFRTVVMPIAHEFSPDVVLVSAGFDAVEGHLSPLGGYSVTARCFGHL

                     1001                                            1050
   HDAC9c    (940)   TKQLMTLADGRVVLALEGGHDLTAICDASEACVNALLGNELEPLAEDILH
 AY032737    (937)   TKQLMTLADGRVVLALEGGHDLTAICDASEACVNALLGNELEPLAEDILH
 AY032738    (880)   --------------------------------------------------
 AF132608    (987)   TRQLMTLAGGRVVLALEGGHDLTAICDASEACVSALLSVELQPLDEAVLQ

                     1051                                            1100
   HDAC9c    (990)   QSPNMNAVISLQKIIEIQSKYWKSVRMVAVPRGCALAGAQLQEETETVSA
 AY032737    (987)   QSPNMNAVISLQKIIEIQSMSLKFS-------------------------
 AY032738    (880)   --------------------------------------------------
 AF132608   (1037)   QKPNINAVATLEKVIEIQSKHWSCVQKFAAGLGRSLREAQAGETEEAETV

                     1101                               1136
   HDAC9c   (1040)   LASLTVDVEQPFAQEDSRTAGEPMEEEPAL------
 AY032737   (1012)   ------------------------------------
 AY032738    (880)   ------------------------------------
 AF132608   (1087)   SAMALESVGAEQAQAAAAREHSPRPAEEPMEQEPAL
```

FIG. 15F

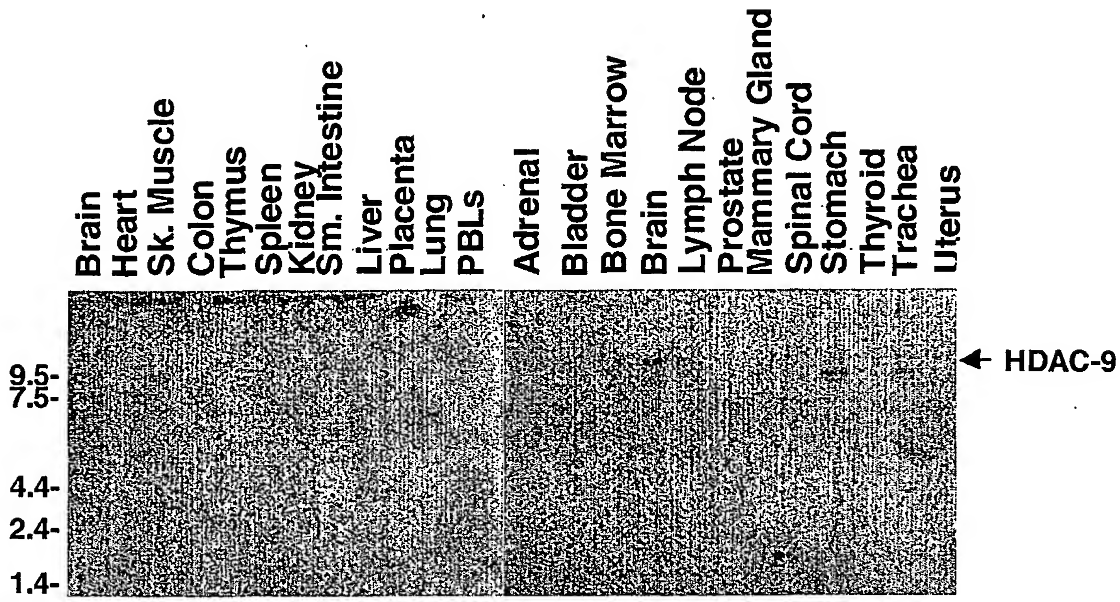


FIG. 16A

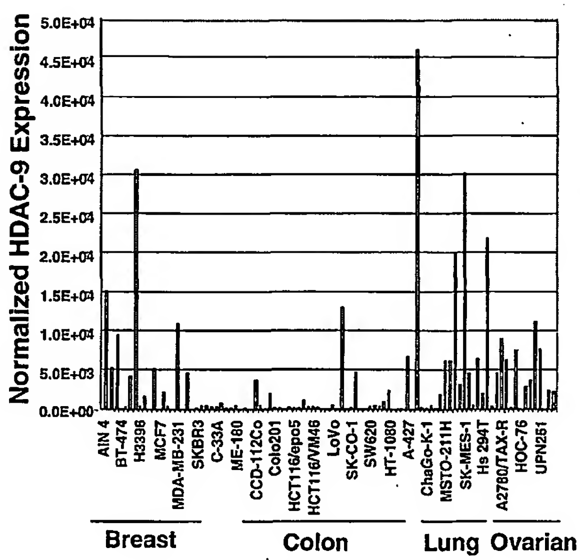


FIG. 16B

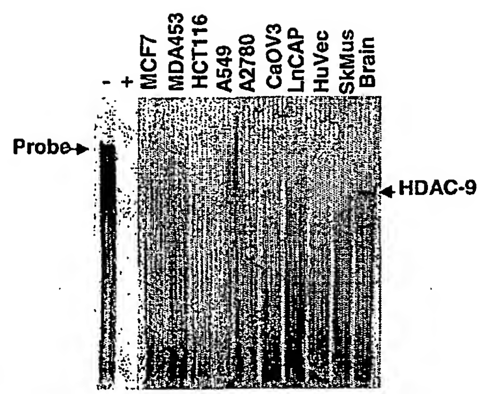


FIG. 16C

```
2901                                          GG AAATGAGCTG GAGCCACTTG
2951  CAGAAGATAT TCTCCACCAA AGCCCGAATA TGAATGCTGT TATTTCTTTA
3001  CAGAAGATCA TTGAAATTCA AAGCAAGTAT TGGAAGTCAG TAAGGATGGT
3051  GGCTGTGCCA AGGGGCTGTG CTCTGGCTGG TGCTCAGTTG CAAGAGGAGA
3101  CAGAGACCGT TTCTGCCCTG GCCTCCCTAA CAGTGGATGT GGAACAGCCC
3151  TTTGCTCAGG AAGACAGCAG AACTGCTGGT GAGCCTATGG AAGAGGAGCC
3201  AGCCTTGTGA
```

FIG. 16D

Tissue Type	Age	Sex	Histology	Surgery	Resected Margin	Stage	HDAC-9/X	b-Actin
Breast	Unk	F	Infiltrating ductal adenocarcinoma	Mastectomy	Positive	2	+4	0
Breast	72	F	Infiltrating ductal adenocarcinoma	Mastectomy	Negative	3	+2	+1
Breast	81	F	Infiltrating ductal adenocarcinoma	Mastectomy	Negative	3	NA	+1
Breast	43	F	Infiltrating ductal adenocarcinoma	Mastectomy		0	+2	+1
Breast	61	F	Infiltrating ductal adenocarcinoma	Mastectomy	Negative	2	+2	+1
Breast	77	F	Infiltrating ductal adenocarcinoma	Mastectomy		3	+2	+1
Breast	69	F	Infiltrating ductal adenocarcinoma	Mastectomy	Positive	3	+3	+1
Breast	76	F	Infiltrating ductal adenocarcinoma	Mastectomy	Negative	2	+2	+1
Breast	Unk	F	Infiltrating ductal adenocarcinoma	Mastectomy	Negative	2	+4	+1
Breast	44	F	Infiltrating ductal adenocarcinoma	Mastectomy		3	+2	0
Breast	61	F	Infiltrating ductal adenocarcinoma	Mastectomy	Negative	2	+2	+1
Breast	46	F	Infiltrating ductal adenocarcinoma	Mastectomy		3	+2	0
Breast	86	F	Infiltrating ductal adenocarcinoma	Biopsy		3	+2	+1
Breast	65	F	Lobular adenocarcinoma	Mastectomy		3	+2	+1
Breast	88	F	Infiltrating ductal adenocarcinoma	Mastectomy		3	+1	0
Breast	47	F	Infiltrating ductal adenocarcinoma	Biopsy		1	+1	+1
Prostate	77	M	Adenocarcinoma	TUR		1	0	+1
Prostate	74	M	Adenocarcinoma	TUR		1	+1	+1
Prostate	55	M	Adenocarcinoma	TUR		1	+1	+1
Prostate	68	M	Adenocarcinoma	TUR		1	+1	+1
Prostate	71	M	Adenocarcinoma	TUR		1	+1	+1
Prostate	66	M	Adenocarcinoma	TUR		1	+2	+1
Prostate	69	M	Adenocarcinoma	TUR		1	+2	+1
Prostate	73	M	Adenocarcinoma	TUR		1	+2	+1
Prostate	72	M	Adenocarcinoma	TUR		1	+1	+1
Prostate	77	M	Adenocarcinoma	TUR		1	+4	+1
Prostate	77	M	Adenocarcinoma	TUR		1	+2	+1
Prostate	73	M	Adenocarcinoma	TUR		1	+2	+1
Prostate	84	M	Adenocarcinoma	TUR		1	+1	+1
Prostate	93	M	Adenocarcinoma	TUR		1	+1	0
Prostate	78	M	Adenocarcinoma	TUR		1	+1	+1
Prostate	78	M	Matched benign specimen	TUR			+1	0

FIG. 17A

Tissue Type	Age	Sex	Histology	Surgery	Resected Margin	Stage	HDAC-9/X	b-Actin
Breast	Unk	F	No pathological diagnosis	Biopsy			0	+1
Breast	Unk	F	No pathological diagnosis	Biopsy			0	+1
Breast	43	F	No pathological diagnosis	Mastectomy			0	0
Breast	88	F	No pathological diagnosis	Mastectomy			0	+1
Breast	55	F	No pathological diagnosis	Mastectomy			0	+1
Breast	74	F	No pathological diagnosis	Mastectomy			+1	+1
Breast	51	F	No pathological diagnosis	Mastectomy			+1	+1
Prostate	69	M	No pathological diagnosis	TUR			0	+1
Prostate	69	M	No pathological diagnosis	TUR			0	+1
Prostate	66	M	No pathological diagnosis	TUR			0	+1
Prostate	69	M	No pathological diagnosis	TUR			0	+1
Prostate	76	M	No pathological diagnosis	TUR			0	+1
Prostate	64	M	No pathological diagnosis	TUR			0	+1
Prostate	66	M	No pathological diagnosis	TUR			0	+1

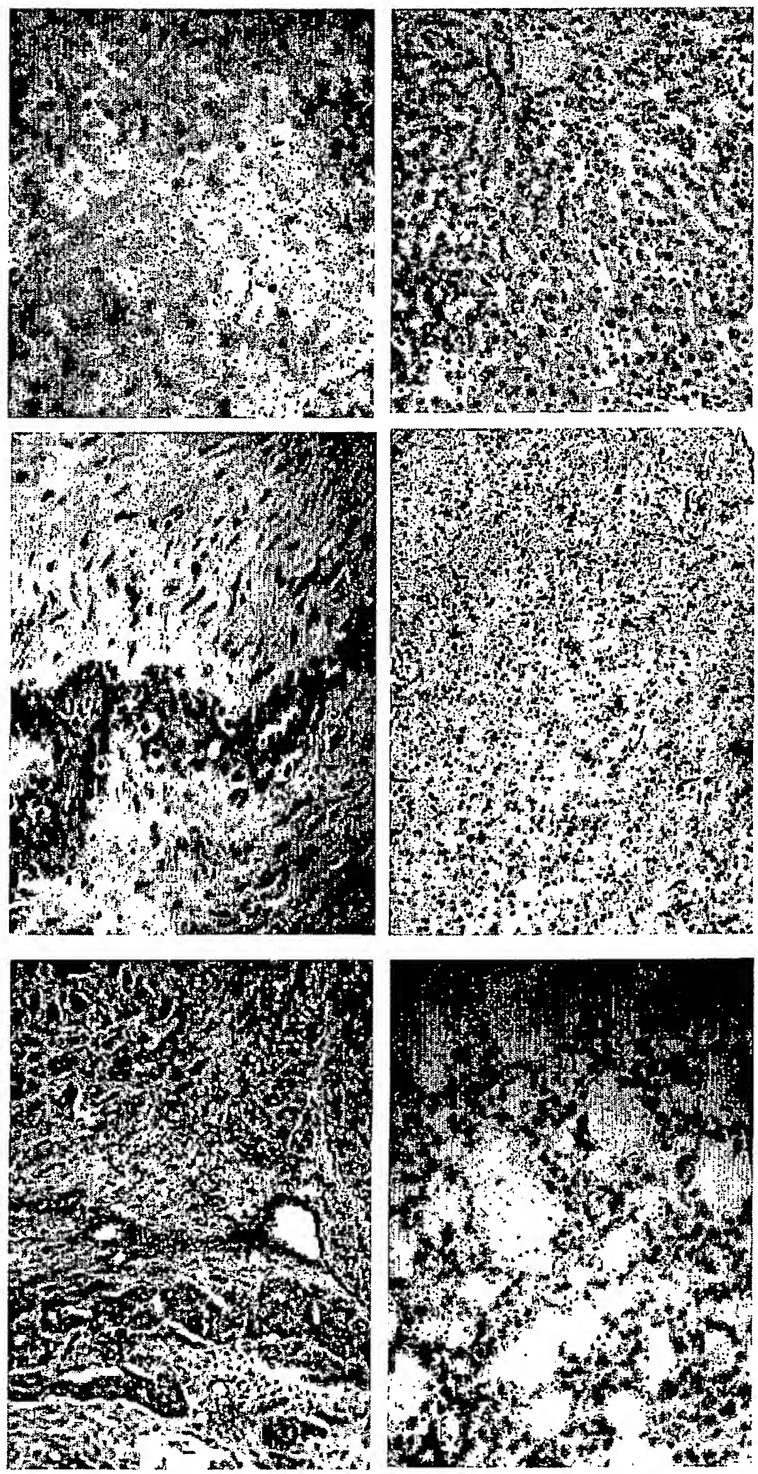
FIG. 17B

FIG. 17C

Breast cancer
Sense, Grade 0

Normal breast
Antisense, Grade 0

Breast cancer
Antisense, Grade 3



Endometrial Cancer
Antisense, Grade 2

Astrocytoma
Antisense, Grade 1

Normal liver
Actin

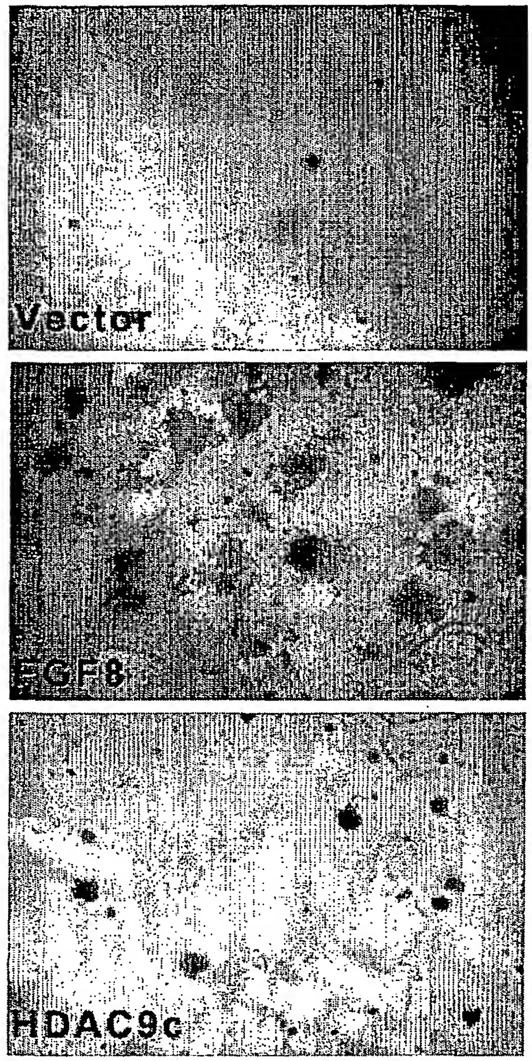


FIG. 18

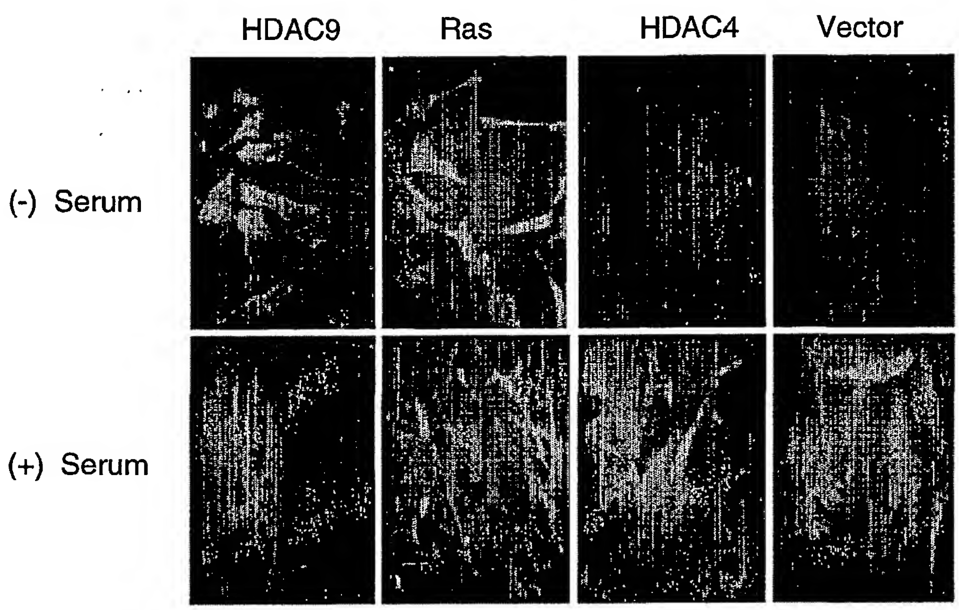


FIG. 19

AlaGluAsnGluThrSerValLeuProProThrProHisAlaGluGlnMetValSerGln
1 GCTGAAAATGAGACTTCGGTTTTGCCCCCTACCCCTCATGCCGAGCAAATGGTTTCACAG

GlnArgIleLeuIleHisGluAspSerMetAsnLeuLeuSerLeuTyrThrSerProSer
61 CAACGCATTCTAATTCATGAAGATTCCATGAACCTGCTAAGTCTTTATACCTCTCCTTCT

LeuProAsnIleThrLeuGlyLeuProAlaValProSerGlnLeuAsnAlaSerAsnSer
121 TTGCCCAACATTACCTTGGGGCTTCCCGCAGTGCCATCCCAGCTCAATGCTTCGAATTCA

LeuLysGluLysGlnLysCysGluThrGlnThrLeuArgGlnGlyValProLeuProGly
181 CTCAAAGAAAAGCAGAAGTGTGAGACGCAGACGCTTAGGCAAGGTGTTCCTCTGCCTGGG

GlnTyrGlyGlySerIleProAlaSerSerSerHisProHisValThrLeuGluGlyLys
241 CAGTATGGAGGCAGCATCCCGGCATCTTCCAGCCACCCTCATGTTACTTTAGAGGGAAAG

ProProAsnSerSerHisGlnAlaLeuLeuGlnHisLeuLeuLeuLysGluGlnMetArg
301 CCACCCAACAGCAGCCACCAGGCTCTCCTGCAGCATTTATTATTGAAAGAACAAATGCGA

GlnGlnLysLeuLeuValAlaGlyGlyValProLeuHisProGlnSerProLeuAlaThr
361 CAGCAAAAGCTTCTTGTAGCTGGTGGAGTTCCCTTACATCCTCAGTCTCCCTTGGCAACA

LysGluArgIleSerProGlyIleArgGlyThrHisLysLeuProArgHisArgProLeu
421 AAAGAGAGAATTTCACCTGGCATTAGAGGTACCCACAAATTGCCCCGTCACAGACCCCTG

AsnArgThrGlnSerAlaProLeuProGlnSerThrLeuAlaGlnLeuValIleGlnGln
481 AACCGAACCCAGTCTGCACCTTTGCCTCAGAGCACGTTGGCTCAGCTGGTCATTCAACAG

GlnHisGlnGlnPheLeuGluLysGlnLysGlnTyrGlnGlnGlnIleHisMetAsnLys
541 CAACACCAGCAATTCTTGGAGAAGCAGAAGCAATACCAGCAGCAGATCCACATGAACAAA

LeuLeuSerLysSerIleGluGlnLeuLysGlnProGlySerHisLeuGluGluAlaGlu
601 CTGCTTTCGAAATCTATTGAACAACTGAAGCAACCAGGCAGTCACCTTGAGGAAGCAGAG

GluGluLeuGlnGlyAspGlnAlaMetGlnGluAspArgAlaProSerSerGlyAsnSer
661 GAAGAGCTTCAGGGGGACCAGGCGATGCAGGAAGACAGAGCGCCCTCTAGTGGCAACAGC

ThrArgSerAspSerSerAlaCysValAspAspThrLeuGlyGlnValGlyAlaValLys
721 ACTAGGAGCGACAGCAGTGCTTGTGTGGATGACACACTGGGACAAGTTGGGGCTGTGAAG

ValLysGluGluProValAspSerAspGluAspAlaGlnIleGlnGluMetGluSerGly
781 GTCAAGGAGGAACCAGTGGACAGTGATGAAGATGCTCAGATCCAGGAAATGGAATCTGGG

GluGlnAlaAlaPheMetGlnGlnProPheLeuGluProThrHisThrArgAlaLeuSer
841 GAGCAGGCTGCTTTTATGCAACAGCCTTTCCTGGAACCCACGCACACACGTGCGCTCTCT

ValArgGlnAlaProLeuAlaAlaValGlyMetAspGlyLeuGluLysHisArgLeuVal
901 GTGCGCCAAGCTCCGCTGGCTGCGGTTGGCATGGATGGATTAGAGAAACACCGTCTCGTC

SerArgThrHisSerSerProAlaAlaSerValLeuProHisProAlaMetAspArgPro
961 TCCAGGACTCACTCTTCCCCTGCTGCCTCTGTTTTACCTCACCCGGCAATGGACCGCCCC

LeuGlnProGlySerAlaThrGlyIleAlaTyrAspProLeuMetLeuLysHisGlnCys
1021 CTCCAGCCTGGCTCTGCAACTGGAATTGCCTATGACCCCTTGATGCTGAAACACCAGTGC

ValCysGlyAsnSerThrThrHisProGluHisAlaGlyArgIleGlnSerIleTrpSer
1081 GTTTGTGGCAATTCCACCACCCACCCTGAGCATGCTGGACGAATACAGAGTATCTGGTCA

FIG. 20A

```
      ArgLeuGlnGluThrGlyLeuLeuAsnLysCysGluArgIleGlnGlyArgLysAlaSer
1141  CGACTGCAAGAAACTGGGCTGCTAAATAAATGTGAGCGAATTCAAGGTCGAAAAGCCAGC

      LeuGluGluIleGlnLeuValHisSerGluHisHisSerLeuLeuTyrGlyThrAsnPro
1201  CTGGAGGAAATACAGCTTGTTCATTCTGAACATCACTCACTGTTGTATGGCACCAACCCC

      LeuAspGlyGlnLysLeuAspProArgIleLeuLeuGlyAspAspSerGlnLysPhePhe
1261  CTGGACGGACAGAAGCTGGACCCCAGGATACTCCTAGGTGATGACTCTCAAAAGTTTTTT

      SerSerLeuProCysGlyGlyLeuGlyValAspSerAspThrIleTrpAsnGluLeuHis
1321  TCCTCATTACCTTGTGGTGGACTTGGGGTGGACAGTGACACCATTTGGAATGAGCTACAC

      SerSerGlyAlaAlaArgMetAlaValGlyCysValIleGluLeuAlaSerLysValAla
1381  TCGTCCGGTGCTGCACGCATGGCTGTTGGCTGTGTCATCGAGCTGGCTTCCAAAGTGGCC

      SerGlyGluLeuLysAsnGlyPheAlaValValArgProProGlyHisHisAlaGluGlu
1441  TCAGGAGAGCTGAAGAATGGGTTTGCTGTTGTGAGGCCCCCTGGCCATCACGCTGAAGAA

      SerThrAlaMetGlyPheCysPhePheAsnSerValAlaIleThrAlaLysTyrLeuArg
1501  TCCACAGCCATGGGGTTCTGCTTTTTTAATTCAGTTGCAATTACCGCCAAATACTTGAGA

      AspGlnLeuAsnIleSerLysIleLeuIleValAspLeuAspValHisHisGlyAsnGly
1561  GACCAACTAAATATAAGCAAGATATTGATTGTAGATCTGGATGTTCACCATGGAAACGGT

      ThrGlnGlnAlaPheTyrAlaAspProSerIleLeuTyrIleSerLeuHisArgTyrAsp
1621  ACCCAGCAGGCCTTTTATGCTGACCCCAGCATCCTGTACATTTCACTCCATCGCTATGAT

      GluGlyAsnPhePheProGlySerGlyAlaProAsnGluValGlyThrGlyLeuGlyGlu
1681  GAAGGGAACTTTTTCCCTGGCAGTGGAGCCCCAAATGAGGTTGGAACAGGCCTTGGAGAA

      GlyTyrAsnIleAsnIleAlaTrpThrGlyGlyLeuAspProProMetGlyAspValGlu
1741  GGGTACAATATAAATATTGCCTGGACAGGTGGCCTTGATCCTCCCATGGGAGATGTTGAG

      TyrLeuGluAlaPheArgThrIleValLysProValAlaLysGluPheAspProAspMet
1801  TACCTTGAAGCATTCAGGACCATCGTGAAGCCTGTGGCCAAAGAGTTTGATCCAGACATG

      ValLeuValSerAlaGlyPheAspAlaLeuGluGlyHisThrProProLeuGlyGlyTyr
1861  GTCTTAGTATCTGCTGGATTTGATGCATTGGAAGGCCACACCCCTCCTCTAGGAGGGTAC

      LysValThrAlaLysCysPheGlyHisLeuThrLysGlnLeuMetThrLeuAlaAspGly
1921  AAAGTGACGGCAAAATGTTTTGGTCATTTGACGAAGCAATTGATGACATTGGCTGATGGA

      ArgValValLeuAlaLeuGluGlyGlyHisAspLeuThrAlaIleCysAspAlaSerGlu
1981  CGTGTGGTGTTGGCTCTAGAAGGAGGACATGATCTCACAGCCATCTGTGATGCATCAGAA

      AlaCysValAsnAlaLeuLeuGlyAsnGluLeuGluProLeuAlaGluAspIleLeuHis
2041  GCCTGTGTAAATGCCCTTCTAGGAAATGAGCTGGAGCCACTTGCAGAAGATATTCTCCAC

      GlnSerProAsnMetAsnAlaValIleSerLeuGlnLysIleIleGluIleGlnSerLys
2101  CAAAGCCCGAATATGAATGCTGTTATTTCTTTACAGAAGATCATTGAAATTCAAAGCAAG

      TyrTrpLysSerValArgMetValAlaValProArgGlyCysAlaLeuAlaGlyAlaGln
2161  TATTGGAAGTCAGTAAGGATGGTGGCTGTGCCAAGGGGCTGTGCTCTGGCTGGTGCTCAG

      LeuGlnGluGluThrGluThrValSerAlaLeuAlaSerLeuThrValAspValGluGln
2221  TTGCAAGAGGAGACAGAGACCGTTTCTGCCCTGGCCTCCCTAACAGTGGATGTGGAACAG
```

FIG. 20B

```
     ProPheAlaGlnGluAspSerArgThrAlaGlyGluProMetGluGluGluProAlaLeu
2281 CCCTTTGCTCAGGAAGACAGCAGAACTGCTGGTGAGCCTATGGAAGAGGAGCCAGCCTTG

     ***
2341 TGAAGTGCCAAGTCCCCCTCTGATATTTCCTGTGTGTGACATCATTGTGTATCCCCCCAC

2401 CCCAGTACCCTCAGACATGTCTTGTCTGCTGCCTGGGTGGCACAGATTCAATGGAACATA
2461 AACACTGGGCACAAAATTCTGAACAGCAGCTTCACTTGTTCTTTGGATGGACTTGAAAGG
2521 GCATTAAAGATTCCTTAAACGTAACCGCTGTGATTCTAGAGTTACAGTAAACCACGATTG
2581 GAAGAAACTGCTTCCAGCATGCTTTTAATATGCTGGGTGACCCACTCCTAGACACCAAGT
2641 TTGAACTAGAAACATTCAGTACAGCACTAGATATTGTTAATTTCAGAAGCTATGACAGCC
2701 AGTGAAATTTTGGGCAAAACCTGAGACATAGTCATTCCTGACATTCTGATCAGCTTTTTT
2761 TGGGGTAATTTGTTTTTCAAACAGTCTTAACTTGTTTACAAGATTTGCTTTTAGCTATGA
2821 ACGGATCGTAATTCCACCCAGAATGTAATGTTTCTTGTTTGTTTGTTTTGTTTTGTTAGG
2881 GTTTTTTTCTCAACTTTAACACACAGTTCAACTGTTCCTAGTAAAAGTTCAAGATGGAGG
2941 AACTAGCATGAGGCTTTTTTCAGTATCTCGAAGTCCAAATGCCAAAGGAACCTCACACAC
3001 TGTTTGTAATGGTGCAATATTTTATATCACTTTTTTTTAAACATCCCCAACATCTTTGTG
3061 TTCTCACACACAGGCAATTTGCAATGTTGCAATTGTGTTGGAGAATGAAGTCCCCCCACC
3121 TCCCAGCCACACACACACATCCTTTGTTCTCATGACAGTAGGTCTGAGCAAATGTTCCACCA
3181 AGCATTTTCAGTGTCTTTGAAAAGCACGTAACTTTTCAAAGGTGGTCTTAATTTGCTGCA
3241 TATCTATCAAGGACTTATTCACTCACCTTTCCTTTTCTGCCCTCTATCAATTGATTTCTT
3301 CTTACCTTTCATCATTCATTCCTTCCTTTAGAAAAACTGAAGATTACCCATAATCTCCTC
3361 TTATTACTTGAGGGCCTTGACTATTTAGTTTATTTTGTTTACTTTACAGGTTAACACAGT
3421 TGTTTTGTCTGATTGCATTTTATTAACTGTGAAGCCGTTGAAATGAATATCACTTAAGCA
3481 ACGTTGCTAAATTTCTATGTGTTTGAAATGTGTTAATGAAGGCACTGCTTATTTGTAGTC
3541 ACCTTGAACTGACTTAACCTAGAAGCTGTGCCTTCTTGTGAAAAAAAAAAAAAAAAAAAA
3601 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

FIG. 20C

```
   1  CCACGCGTCCGTAGGAGAAGGGCACCGGCTGGAGCCACTTGCAGGACTGAGGGTTTTTGC
  61  AACAAAACCCTAGCAGCCTGAAGAACTCTAAGCCAGGTTTAATTGGTTTCTTTTTCTCGT
 121  GGGTAGACTTAATAATTTTCTACGTATTCTGACAAAGAAATAACCCCGAAGCACGTTCCT
 181  ATTTCCCACCTGCTTGTAGTTTCCGGGATAACCTAAACTCCAGAGAGCTATAGCATCCAC
 241  TCTGTCCTTTCTGCTTTGCACACAGATGGGGTGGCTGGACGAGAGCAGCTCTTGGCTCAG

              MetHisSerMetIleSerSerValAspValLysSerGluValProValGlyLeu
 301  CAAAGAATGCACAGTATGATCAGCTCAGTGGATGTGAAGTCAGAAGTTCCTGTGGGCCTG

      GluProIleSerProLeuAspLeuArgThrAspLeuArgMetMetMetProValValAsp
 361  GAGCCCATCTCACCTTTAGACCTAAGGACAGACCTCAGGATGATGATGCCCGTGGTGGAC

      ProValValArgGluLysGlnLeuGlnGluLeuLeuLeuIleGlnGlnGlnGlnGln
 421  CCTGTTGTCCGTGAGAAGCAATTGCAGCAGGAATTACTTCTTATCCAGCAGCAGCAACAA

      IleGlnLysGlnLeuLeuIleAlaGluPheGlnLysGlnHisGluAsnLeuThrArgGln
 481  ATCCAGAAGCAGCTTCTGATAGCAGAGTTTCAGAAACAGCATGAGAACTTGACACGGCAG

      HisGlnAlaGlnLeuGlnGluHisIleLysLeuGlnGlnGluLeuLeuAlaIleLysGln
 541  CACCAGGCTCAGCTTCAGGAGCATATCAAGTTGCAACAGGAACTTCTAGCCATAAAACAG

      GlnGlnGluLeuLeuGluLysGluGlnLysLeuGluGlnGlnArgGlnGluGlnGluVal
 601  CAACAAGAACTCCTAGAAAAGGAGCAGAAACTGGAGCAGCAGAGGCAAGAACAGGAAGTA

      GluArgHisArgArgGluGlnGlnLeuProProLeuArgGlyLysAspArgGlyArgGlu
 661  GAGAGGCATCGCAGAGAACAGCAGCTTCCTCCTCTCAGAGGCAAAGATAGAGGACGAGAA

      ArgAlaValAlaSerThrGluValLysGlnLysLeuGlnGluPheLeuLeuSerLysSer
 721  AGGGCAGTGGCAAGTACAGAAGTAAAGCAGAAGCTTCAAGAGTTCCTACTGAGTAAATCA

      AlaThrLysAspThrProThrAsnGlyLysAsnHisSerValSerArgHisProLysLeu
 781  GCAACGAAAGACACTCCAACTAATGGAAAAAATCATTCCGTGAGCCGCCATCCCAAGCTC

      TrpTyrThrAlaAlaHisHisThrSerLeuAspGlnSerSerProProLeuSerGlyThr
 841  TGGTACACGGCTGCCCACCACACATCATTGGATCAAAGCTCTCCACCCCTTAGTGGAACA

      SerProSerTyrLysTyrThrLeuProGlyAlaGlnAspAlaLysAspAspPheProLeu
 901  TCTCCATCCTACAAGTACACATTACCAGGAGCACAAGATGCAAAGGATGATTTCCCCCTT

      ArgLysThrAlaSerGluProAsnLeuLysValArgSerArgLeuLysGlnLysValAla
 961  CGAAAAACTGCCTCTGAGCCCAACTTGAAGGTGCGGTCCAGGTTAAAACAGAAAGTGGCA

      GluArgArgSerSerProLeuLeuArgArgLysAspGlyAsnValValThrSerPheLys
1021  GAGAGGAGAAGCAGCCCCTTACTCAGGCGGAAGGATGGAAATGTTGTCACTTCATTCAAG

      LysArgMetPheGluValThrGluSerSerValSerSerSerSerProGlySerGlyPro
1081  AAGCGAATGTTTGAGGTGACAGAATCCTCAGTCAGTAGCAGTTCTCCAGGCTCTGGTCCC

      SerSerProAsnAsnGlyProThrGlySerValThrGluAsnGluThrSerValLeuPro
1141  AGTTCACCAAACAATGGGCCAACTGGAAGTGTTACTGAAAATGAGACTTCGGTTTTGCCC

      ProThrProHisAlaGluGlnMetValSerGlnGlnArgIleLeuIleHisGluAspSer
1201  CCTACCCCTCATGCCGAGCAAATGGTTTCACAGCAACGCATTCTAATTCATGAAGATTCC

      MetAsnLeuLeuSerLeuTyrThrSerProSerLeuProAsnIleThrLeuGlyLeuPro
1261  ATGAACCTGCTAAGTCTTTATACCTCTCCTTCTTTGCCCAACATTACCTTGGGGCTTCCC
```

FIG. 21A

```
      AlaValProSerGlnLeuAsnAlaSerAsnSerLeuLysGluLysGlnLysCysGluThr
1321  GCAGTGCCATCCCAGCTCAATGCTTCGAATTCACTCAAAGAAAAGCAGAAGTGTGAGACG

      GlnThrLeuArgGlnGlyValProLeuProGlyGlnTyrGlyGlySerIleProAlaSer
1381  CAGACGCTTAGGCAAGGTGTTCCTCTGCCTGGGCAGTATGGAGGCAGCATCCCGGCATCT

      SerSerHisProHisValThrLeuGluGlyLysProProAsnSerSerHisGlnAlaLeu
1441  TCCAGCCACCCTCATGTTACTTTAGAGGGAAAGCCACCCAACAGCAGCCACCAGGCTCTC

      LeuGlnHisLeuLeuLeuLysGluGlnMetArgGlnGlnLysLeuLeuValAlaGlyGly
1501  CTGCAGCATTTATTATTGAAAGAACAAATGCGACAGCAAAAGCTTCTTGTAGCTGGTGGA

      ValProLeuHisProGlnSerProLeuAlaThrLysGluArgIleSerProGlyIleArg
1561  GTTCCCTTACATCCTCAGTCTCCCTTGGCAACAAAAGAGAGAATTTCACCTGGCATTAGA

      GlyThrHisLysLeuProArgHisArgProLeuAsnArgThrGlnSerAlaProLeuPro
1621  GGTACCCACAAATTGCCCCGTCACAGACCCCTGAACCGAACCCAGTCTGCACCTTTGCCT

      GlnSerThrLeuAlaGlnLeuValIleGlnGlnGlnHisGlnGlnPheLeuGluLysGln
1681  CAGAGCACGTTGGCTCAGCTGGTCATTCAACAGCAACACCAGCAATTCTTGGAGAAGCAG

      LysGlnTyrGlnGlnGlnIleHisMetAsnLysGluLeuProMetThrPro***
1741  AAGCAATACCAGCAGCAGATCCACATGAACAAAGAATTGCCTATGACCCCTTGATGCTGA

1801  AACACCAGTGCGTTTGTGGCAATTCCACCACCCACCCTGAGCATGCTGGACGAATACAGA
1861  GTATCTGGTCACGACTGCAAGAAACTGGGCTGCTAAATAAATGTGAGCGAATTCAAGGTC
1921  GAAAAGCCAGCCTGGAGGAAATACAGCTTGTTCATTCTGAACATCACTCACTGTTGTATG
1981  GCACCAACCCCCTGGACGGACAGAAGCTGGACCCCAGGATACTCCTAGGTGATGACTCTC
2041  AAAAGTTTTTTTCCTCATTACCTTGTGGTGGACTTGGGGTGGACAGTGACACCATTTGGA
2101  ATGAGCTACACTCGTCCGGTGCTGCACGCATGGCTGTTGGCTGTGTCATCGAGCTGGCTT
2161  CCAAAGTGGCCTCAGGAGAGCTGAAGGTGAGGTCCGGGTTGCATTAAGTGTGGGAAATCC
2221  AGAGAAGAAACTGAAACAGAGATGTTGTTATGTGGGAATTGCGGGGAGTGTGGCGTGGTA
2281  ATAAAAGGAAGGGCAGAAGGAAGAGGGTAGAGATGGCCACTAAGGTGTGATAATAACTCA
2341  TCTGTAGGCAGGGAGCAGCTCATCCTGCTCTCAGGGCCTTCTTCTGCCTGAGAACACTCT
2401  GCAGTCAGGGCCCACCGGTGTGCATGTAAGAGCACAGAGATAATAAGCAAAGCTATGGTT
2461  CAGGTTAAAAATACCTTTAGTATATACATGTCTGTCATGCCATCCTGAGATTCTCTTTTG
2521  AGGCAATTTTAAAAATATGATTACTGAGAAGTGTGTATAAGCTCAGAATACCACCCAGAG
2581  AGAGGGAGGCAGAGAAAGGTAAATACCAGACGGGAAGGATTGGGAGGAGGAAGGAAATTG
2641  TTGATTAGAAGGGTAATGATCCAGAGTGTGTTTTTCCATGAAAGAACTTAAAAAATGAGC
2701  TATGCTTTATTGTTCTTTTCTTTTTATGGTCTCTTCTTTTCTACATCGTATGAAAAGAAC
2761  AATGTCCAAACCCCAGCGTTTCCCAGTCTAAACAATTTATAAAAGCTAGAGACCTGACAG
2821  ACGTTGACATTTTATTTGGTATTTTAACAGTGCTATTTAAAGGTACGCCATGTGCGTCTT
2881  GAATGCAGTTACCCCAATAAACTTTGTTGGTGCTAACACGGCCTTTTAATGCACTAGTTC
2941  ACACACTTCATGACGCAATCTGGGTCGTGATTGATTCGGTATTTTTAGCAATTGCGGGGC
3001  TTAGGGAAATATATTATGACCAATAACATATGCACTGTGAGTTTTGTGAAACCAAGATAA
3061  AATAATTAGGATTACTTTTCTTTATGTCTAGTGAATTTTTATTCAATTACATGGGACTCT
3121  TCCAGTTGTGATTAAAAATGTGGAGTAGGAATGTGCACTTCACAATGCAACGTTTGTCCA
3181  AGAAGTCTTTACTCTTAACTCTTTAAAGAGTCAGAGCCTACGGAAATATAATTTTGATAG
3241  GGTGAGCTCTATTTAAAAAGTAGATGTGCCTGTATATATTTGACATAAGTAGTATTAGGA
3301  CATTGCTCATCTCAGGGGATATATGGGGTCATTAATGTGGTGCTTACTCTTCAGTCTTTA
3361  CCTTTGAAAATGAGCAAAAAAAAAAAAAAAA
```

FIG. 21B

1 GGGGAAGAGAGGCACAGACACAGATAGGAGAAGGGCACCGGCTGGAGCCACTTGCAGGAC
61 TGAGGGTTTTTGCAACAAAACCCTAGCAGCCTGAAGAACTCTAAGCCAGATGGGGTGGCT

MetHisSerMetIleSerSerValAspVal
121 GGACGAGAGCAGCTCTTGGCTCAGCAAAGAATGCACAGTATGATCAGCTCAGTGGATGTG

LysSerGluValProValGlyLeuGluProIleSerProLeuAspLeuArgThrAspLeu
181 AAGTCAGAAGTTCCTGTGGGCCTGGAGCCCATCTCACCTTTAGACCTAAGGACAGACCTC

ArgMetMetMetProValValAspProValValArgGluLysGlnLeuGlnGlnGluLeu
241 AGGATGATGATGCCCGTGGTGGACCCTGTTGTCCGTGAGAAGCAATTGCAGCAGGAATTA

LeuLeuIleGlnGlnGlnGlnGlnIleGlnLysGlnLeuLeuIleAlaGluPheGlnLys
301 CTTCTTATCCAGCAGCAGCAACAAATCCAGAAGCAGCTTCTGATAGCAGAGTTTCAGAAA

GlnHisGluAsnLeuThrArgGlnHisGlnAlaGlnLeuGlnGluHisIleLysGluLeu
361 CAGCATGAGAACTTGACACGGCAGCACCAGGCTCAGCTTCAGGAGCATATCAAGGAACTT

LeuAlaIleLysGlnGlnGlnGluLeuLeuGluLysGluGlnLysLeuGluGlnGlnArg
421 CTAGCCATAAAACAGCAACAAGAACTCCTAGAAAAGGAGCAGAAACTGGAGCAGCAGAGG

GlnGluGlnGluValGluArgHisArgArgGluGlnGlnLeuProProLeuArgGlyLys
481 CAAGAACAGGAAGTAGAGAGGCATCGCAGAGAACAGCAGCTTCCTCCTCTCAGAGGCAAA

AspArgGlyArgGluArgAlaValAlaSerThrGluValLysGlnLysLeuGlnGluPhe
541 GATAGAGGACGAGAAAGGGCAGTGGCAAGTACAGAAGTAAAGCAGAAGCTTCAAGAGTTC

LeuLeuSerLysSerAlaThrLysAspThrProThrAsnGlyLysAsnHisSerValSer
601 CTACTGAGTAAATCAGCAACGAAAGACACTCCAACTAATGGAAAAAATCATTCCGTGAGC

ArgHisProLysLeuTrpTyrThrAlaAlaHisHisThrSerLeuAspGlnSerSerPro
661 CGCCATCCCAAGCTCTGGTACACGGCTGCCCACCACACATCATTGGATCAAAGCTCTCCA

ProLeuSerGlyThrSerProSerTyrLysTyrThrLeuProGlyAlaGlnAspAlaLys
721 CCCCTTAGTGGAACATCTCCATCCTACAAGTACACATTACCAGGAGCACAAGATGCAAAG

AspAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArgLeu
781 GATGATTTCCCCCTTCGAAAAACTGCCTCTGAGCCCAACTTGAAGGTGCGGTCCAGGTTA

LysGlnLysValAlaGluArgArgSerSerProLeuLeuArgArgLysAspGlyAsnVal
841 AAACAGAAAGTGGCAGAGAGGAGAAGCAGCCCCTTACTCAGGCGGAAGGATGGAAATGTT

ValThrSerPheLysLysArgMetPheGluValThrGluSerSerValSerSerSerSer
901 GTCACTTCATTCAAGAAGCGAATGTTTGAGGTGACAGAATCCTCAGTCAGTAGCAGTTCT

ProGlySerGlyProSerSerProAsnAsnGlyProThrGlySerValThrGluAsnGlu
961 CCAGGCTCTGGTCCCAGTTCACCAAACAATGGGCCAACTGGAAGTGTTACTGAAAATGAG

ThrSerValLeuProProThrProHisAlaGluGlnMetValSerGlnGlnArgIleLeu
1021 ACTTCGGTTTTGCCCCCTACCCCTCATGCCGAGCAAATGGTTTCACAGCAACGCATTCTA

IleHisGluAspSerMetAsnLeuLeuSerLeuTyrThrSerProSerLeuProAsnIle
1081 ATTCATGAAGATTCCATGAACCTGCTAAGTCTTTATACCTCTCCTTCTTTGCCCAACATT

ThrLeuGlyLeuProAlaValProSerGlnLeuAsnAlaSerAsnSerLeuLysGluLys
1141 ACCTTGGGGCTTCCCGCAGTGCCATCCCAGCTCAATGCTTCGAATTCACTCAAAGAAAAG

FIG. 22A

```
      GlnLysCysGluThrGlnThrLeuArgGlnGlyValProLeuProGlyGlnTyrGlyGly
1201  CAGAAGTGTGAGACGCAGACGCTTAGGCAAGGTGTTCCTCTGCCTGGGCAGTATGGAGGC

      SerIleProAlaSerSerSerHisProHisValThrLeuGluGlyLysProProAsnSer
1261  AGCATCCCGGCATCTTCCAGCCACCCTCATGTTACTTTAGAGGGAAAGCCACCCAACAGC

      SerHisGlnAlaLeuLeuGlnHisLeuLeuLeuLysGluGlnMetArgGlnGlnLysLeu
1321  AGCCACCAGGCTCTCCTGCAGCATTTATTATTGAAAGAACAAATGCGACAGCAAAAGCTT

      LeuValAlaGlyGlyValProLeuHisProGlnSerProLeuAlaThrLysGluArgIle
1381  CTTGTAGCTGGTGGAGTTCCCTTACATCCTCAGTCTCCCTTGGCAACAAAAGAGAGAATT

      SerProGlyIleArgGlyThrHisLysLeuProArgHisArgProLeuAsnArgThrGln
1441  TCACCTGGCATTAGAGGTACCCACAAATTGCCCCGTCACAGACCCCTGAACCGAACCCAG

      SerAlaProLeuProGlnSerThrLeuAlaGlnLeuValIleGlnGlnGlnHisGlnGln
1501  TCTGCACCTTTGCCTCAGAGCACGTTGGCTCAGCTGGTCATTCAACAGCAACACCAGCAA

      PheLeuGluLysGlnLysGlnTyrGlnGlnGlnIleHisMetAsnLysLeuLeuSerLys
1561  TTCTTGGAGAAGCAGAAGCAATACCAGCAGCAGATCCACATGAACAAACTGCTTTCGAAA

      SerIleGluGlnLeuLysGlnProGlySerHisLeuGluGluAlaGluGluGluLeuGln
1621  TCTATTGAACAACTGAAGCAACCAGGCAGTCACCTTGAGGAAGCAGAGGAAGAGCTTCAG

      GlyAspGlnAlaMetGlnGluAspArgAlaProSerSerGlyAsnSerThrArgSerAsp
1681  GGGGACCAGGCGATGCAGGAAGACAGAGCGCCCTCTAGTGGCAACAGCACTAGGAGCGAC

      SerSerAlaCysValAspAspThrLeuGlyGlnValGlyAlaValLysValLysGluGlu
1741  AGCAGTGCTTGTGTGGATGACACACTGGGACAAGTTGGGGCTGTGAAGGTCAAGGAGGAA

      ProValAspSerAspGluAspAlaGlnIleGlnGluMetGluSerGlyGluGlnAlaAla
1801  CCAGTGGACAGTGATGAAGATGCTCAGATCCAGGAAATGGAATCTGGGGAGCAGGCTGCT

      PheMetGlnGlnProPheLeuGluProThrHisThrArgAlaLeuSerValArgGlnAla
1861  TTTATGCAACAGCCTTTCCTGGAACCCACGCACACACGTGCGCTCTCTGTGCGCCAAGCT

      ProLeuAlaAlaValGlyMetAspGlyLeuGluLysHisArgLeuValSerArgThrHis
1921  CCGCTGGCTGCGGTTGGCATGGATGGATTAGAGAAACACCGTCTCGTCTCCAGGACTCAC

      SerSerProAlaAlaSerValLeuProHisProAlaMetAspArgProLeuGlnProGly
1981  TCTTCCCCTGCTGCCTCTGTTTTACCTCACCCAGCAATGGACCGCCCCCTCCAGCCTGGC

      SerAlaThrGlyIleAlaTyrAspProLeuMetLeuLysHisGlnCysValCysGlyAsn
2041  TCTGCAACTGGAATTGCCTATGACCCCTTGATGCTGAAACACCAGTGCGTTTGTGGCAAT

      SerThrThrHisProGluHisAlaGlyArgIleGlnSerIleTrpSerArgLeuGlnGlu
2101  TCCACCACCCACCCTGAGCATGCTGGACGAATACAGAGTATCTGGTCACGACTGCAAGAA

      ThrGlyLeuLeuAsnLysCysGluArgIleGlnGlyArgLysAlaSerLeuGluGluIle
2161  ACTGGGCTGCTAAATAAATGTGAGCGAATTCAAGGTCGAAAAGCCAGCCTGGAGGAAATA

      GlnLeuValHisSerGluHisHisSerLeuLeuTyrGlyThrAsnProLeuAspGlyGln
2221  CAGCTTGTTCATTCTGAACATCACTCACTGTTGTATGGCACCAACCCCCTGGACGGACAG

      LysLeuAspProArgIleLeuLeuGlyAspAspSerGlnLysPhePheSerSerLeuPro
2281  AAGCTGGACCCCAGGATACTCCTAGGTGATGACTCTCAAAAGTTTTTTTCCTCATTACCT
```

FIG. 22B

```
      CysGlyGlyLeuGlyValAspSerAspThrIleTrpAsnGluLeuHisSerSerGlyAla
2341  TGTGGTGGACTTGGGGTGGACAGTGACACCATTTGGAATGAGCTACACTCGTCCGGTGCT

      AlaArgMetAlaValGlyCysValIleGluLeuAlaSerLysValAlaSerGlyGluLeu
2401  GCACGCATGGCTGTTGGCTGTGTCATCGAGCTGGCTTCCAAAGTGGCCTCAGGAGAGCTG

      LysAsnGlyPheAlaValValArgProProGlyHisHisAlaGluGluSerThrAlaMet
2461  AAGAATGGGTTTGCTGTTGTGAGGCCCCCTGGCCATCACGCTGAAGAATCCACAGCCATG

      GlyPheCysPhePheAsnSerValAlaIleThrAlaLysTyrLeuArgAspGlnLeuAsn
2521  GGGTTCTGCTTTTTTAATTCAGTTGCAATTACCGCCAAATACTTGAGAGACCAACTAAAT

      IleSerLysIleLeuIleValAspLeuAspValHisHisGlyAsnGlyThrGlnGlnAla
2581  ATAAGCAAGATATTGATTGTAGATCTGGATGTTCACCATGGAAACGGTACCCAGCAGGCC

      PheTyrAlaAspProSerIleLeuTyrIleSerLeuHisArgTyrAspGluGlyAsnPhe
2641  TTTTATGCTGACCCCAGCATCCTGTACATTTCACTCCATCGCTATGATGAAGGGAACTTT

      ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      PheProGlySerGlyAlaProAsnGluValGlyThrGlyLeuGlyGluGlyTyrAsnIle
2701  TTCCCTGGCAGTGGAGCCCCAAATGAGGTTGGAACAGGCCTTGGAGAAGGGTACAATATA

      AsnIleAlaTrpThrGlyGlyLeuAspProProMetGlyAspValGluTyrLeuGluAla
2761  AATATTGCCTGGACAGGTGGCCTTGATCCTCCCATGGGAGATGTTGAGTACCTTGAAGCA

      PheArgThrIleValLysProValAlaLysGluPheAspProAspMetValLeuValSer
2821  TTCAGGACCATCGTGAAGCCTGTGGCCAAAGAGTTTGATCCAGACATGGTCTTAGTATCT

      AlaGlyPheAspAlaLeuGluGlyHisThrProProLeuGlyGlyTyrLysValThrAla
2881  GCTGGATTTGATGCATTGGAAGGCCACACCCCTCCTCTAGGAGGGTACAAAGTGACGGCA

      LysCysPheGlyHisLeuThrLysGlnLeuMetThrLeuAlaAspGlyArgValValLeu
2941  AAATGTTTTGGTCATTTGACGAAGCAATTGATGACATTGGCTGATGGACGTGTGGTGTTG

      AlaLeuGluGlyGlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaCysValAsn
3001  GCTCTAGAAGGAGGACATGATCTCACAGCCATCTGTGATGCATCAGAAGCCTGTGTAAAT

      AlaLeuLeuGlyAsnGluLeuGluProLeuAlaGluAspIleLeuHisGlnSerProAsn
3061  GCCCTTCTAGGAAATGAGCTGGAGCCACTTGCAGAAGATATTCTCCACCAAAGCCCGAAT

      MetAsnAlaValIleSerLeuGlnLysIleIleGluIleGlnSerMetSerLeuLysPhe
3121  ATGAATGCTGTTATTTCTTTACAGAAGATCATTGAAATTCAAAGTATGTCTTTAAAGTTC

      Ser***
3181  TCTTAA
```

FIG. 22C

```
   1  GGGGAAGAGAGGCACAGACACAGATAGGAGAAGGGCACCGGCTGGAGCCACTTGCAGGAC
  61  TGAGGGTTTTTGCAACAAAACCCTAGCAGCCTGAAGAACTCTAAGCCAGATGGGGTGGCT

                                              MetHisSerMetIleSerSerValAspVal
 121  GGACGAGAGCAGCTCTTGGCTCAGCAAAGAATGCACAGTATGATCAGCTCAGTGGATGTG

      LysSerGluValProValGlyLeuGluProIleSerProLeuAspLeuArgThrAspLeu
 181  AAGTCAGAAGTTCCTGTGGGCCTGGAGCCCATCTCACCTTTAGACCTAAGGACAGACCTC

      ArgMetMetMetProValValAspProValValArgGluLysGlnLeuGlnGlnGluLeu
 241  AGGATGATGATGCCCGTGGTGGACCCTGTTGTCCGTGAGAAGCAATTGCAGCAGGAATTA

      LeuLeuIleGlnGlnGlnGlnGlnIleGlnLysGlnLeuLeuIleAlaGluPheGlnLys
 301  CTTCTTATCCAGCAGCAGCAACAAATCCAGAAGCAGCTTCTGATAGCAGAGTTTCAGAAA

      GlnHisGluAsnLeuThrArgGlnHisGlnAlaGlnLeuGlnGluHisIleLysGluLeu
 361  CAGCATGAGAACTTGACACGGCAGCACCAGGCTCAGCTTCAGGAGCATATCAAGGAACTT

      LeuAlaIleLysGlnGlnGlnGluLeuLeuGluLysGluGlnLysLeuGluGlnGlnArg
 421  CTAGCCATAAAACAGCAACAAGAACTCCTAGAAAAGGAGCAGAAACTGGAGCAGCAGAGG

      GlnGluGlnGluValGluArgHisArgArgGluGlnGlnLeuProProLeuArgGlyLys
 481  CAAGAACAGGAAGTAGAGAGGCATCGCAGAGAACAGCAGCTTCCTCCTCTCAGAGGCAAA

      AspArgGlyArgGluArgAlaValAlaSerThrGluValLysGlnLysLeuGlnGluPhe
 541  GATAGAGGACGAGAAAGGGCAGTGGCAAGTACAGAAGTAAAGCAGAAGCTTCAAGAGTTC

      LeuLeuSerLysSerAlaThrLysAspThrProThrAsnGlyLysAsnHisSerValSer
 601  CTACTGAGTAAATCAGCAACGAAAGACACTCCAACTAATGGAAAAAATCATTCCGTGAGC

      ArgHisProLysLeuTrpTyrThrAlaAlaHisHisThrSerLeuAspGlnSerSerPro
 661  CGCCATCCCAAGCTCTGGTACACGGCTGCCCACCACACATCATTGGATCAAAGCTCTCCA

      ProLeuSerGlyThrSerProSerTyrLysTyrThrLeuProGlyAlaGlnAspAlaLys
 721  CCCCTTAGTGGAACATCTCCATCCTACAAGTACACATTACCAGGAGCACAAGATGCAAAG

      AspAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArgLeu
 781  GATGATTTCCCCCTTCGAAAAACTGCCTCTGAGCCCAACTTGAAGGTGCGGTCCAGGTTA

      LysGlnLysValAlaGluArgArgSerSerProLeuLeuArgArgLysAspGlyAsnVal
 841  AAACAGAAAGTGGCAGAGAGGAGAAGCAGCCCCTTACTCAGGCGGAAGGATGGAAATGTT

      ValThrSerPheLysLysArgMetPheGluValThrGluSerSerValSerSerSerSer
 901  GTCACTTCATTCAAGAAGCGAATGTTTGAGGTGACAGAATCCTCAGTCAGTAGCAGTTCT

      ProGlySerGlyProSerSerProAsnAsnGlyProThrGlySerValThrGluAsnGlu
 961  CCAGGCTCTGGTCCCAGTTCACCAAACAATGGGCCAACTGGAAGTGTTACTGAAAATGAG

      ThrSerValLeuProProThrProHisAlaGluGlnMetValSerGlnGlnArgIleLeu
1021  ACTTCGGTTTTGCCCCCTACCCCTCATGCCGAGCAAATGGTTTCACAGCAACGCATTCTA

      IleHisGluAspSerMetAsnLeuLeuSerLeuTyrThrSerProSerLeuProAsnIle
1081  ATTCATGAAGATTCCATGAACCTGCTAAGTCTTTATACCTCTCCTTCTTTGCCCAACATT

      ThrLeuGlyLeuProAlaValProSerGlnLeuAsnAlaSerAsnSerLeuLysGluLys
1141  ACCTTGGGGCTTCCCGCAGTGCCATCCCAGCTCAATGCTTCGAATTCACTCAAAGAAAAG
```

FIG. 22D

GlnLysCysGluThrGlnThrLeuArgGlnGlyValProLeuProGlyGlnTyrGlyGly
1201 CAGAAGTGTGAGACGCAGACGCTTAGGCAAGGTGTTCCTCTGCCTGGGCAGTATGGAGGC

SerIleProAlaSerSerSerHisProHisValThrLeuGluGlyLysProProAsnSer
1261 AGCATCCCGGCATCTTCCAGCCACCCTCATGTTACTTTAGAGGGAAAGCCACCCAACAGC

SerHisGlnAlaLeuLeuGlnHisLeuLeuLeuLysGluGlnMetArgGlnGlnLysLeu
1321 AGCCACCAGGCTCTCCTGCAGCATTTATTATTGAAAGAACAAATGCGACAGCAAAAGCTT

LeuValAlaGlyGlyValProLeuHisProGlnSerProLeuAlaThrLysGluArgIle
1381 CTTGTAGCTGGTGGAGTTCCCTTACATCCTCAGTCTCCCTTGGCAACAAAAGAGAGAATT

SerProGlyIleArgGlyThrHisLysLeuProArgHisArgProLeuAsnArgThrGln
1441 TCACCTGGCATTAGAGGTACCCACAAATTGCCCCGTCACAGACCCCTGAACCGAACCCAG

SerAlaProLeuProGlnSerThrLeuAlaGlnLeuValIleGlnGlnGlnHisGlnGln
1501 TCTGCACCTTTGCCTCAGAGCACGTTGGCTCAGCTGGTCATTCAACAGCAACACCAGCAA

PheLeuGluLysGlnLysGlnTyrGlnGlnGlnIleHisMetAsnLysLeuLeuSerLys
1561 TTCTTGGAGAAGCAGAAGCAATACCAGCAGCAGATCCACATGAACAAACTGCTTTCGAAA

SerIleGluGlnLeuLysGlnProGlySerHisLeuGluGluAlaGluGluGluLeuGln
1621 TCTATTGAACAACTGAAGCAACCAGGCAGTCACCTTGAGGAAGCAGAGGAAGAGCTTCAG

GlyAspGlnAlaMetGlnGluAspArgAlaProSerSerGlyAsnSerThrArgSerAsp
1681 GGGGACCAGGCGATGCAGGAAGACAGAGCGCCCTCTAGTGGCAACAGCACTAGGAGCGAC

SerSerAlaCysValAspAspThrLeuGlyGlnValGlyAlaValLysValLysGluGlu
1741 AGCAGTGCTTGTGTGGATGACACACTGGGACAAGTTGGGGCTGTGAAGGTCAAGGAGGAA

ProValAspSerAspGluAspAlaGlnIleGlnGluMetGluSerGlyGluGlnAlaAla
1801 CCAGTGGACAGTGATGAAGATGCTCAGATCCAGGAAATGGAATCTGGGGAGCAGGCTGCT

PheMetGlnGlnProPheLeuGluProThrHisThrArgAlaLeuSerValArgGlnAla
1861 TTTATGCAACAGCCTTTCCTGGAACCCACGCACACACGTGCGCTCTCTGTGCGCCAAGCT

ProLeuAlaAlaValGlyMetAspGlyLeuGluLysHisArgLeuValSerArgThrHis
1921 CCGCTGGCTGCGGTTGGCATGGATGGATTAGAGAAACACCGTCTCGTCTCCAGGACTCAC

SerSerProAlaAlaSerValLeuProHisProAlaMetAspArgProLeuGlnProGly
1981 TCTTCCCCTGCTGCCTCTGTTTTACCTCACCCAGCAATGGACCGCCCCCTCCAGCCTGGC

SerAlaThrGlyIleAlaTyrAspProLeuMetLeuLysHisGlnCysValCysGlyAsn
2041 TCTGCAACTGGAATTGCCTATGACCCCTTGATGCTGAAACACCAGTGCGTTTGTGGCAAT

SerThrThrHisProGluHisAlaGlyArgIleGlnSerIleTrpSerArgLeuGlnGlu
2101 TCCACCACCCACCCTGAGCATGCTGGACGAATACAGAGTATCTGGTCACGACTGCAAGAA

ThrGlyLeuLeuAsnLysCysGluArgIleGlnGlyArgLysAlaSerLeuGluGluIle
2161 ACTGGGCTGCTAAATAAATGTGAGCGAATTCAAGGTCGAAAAGCCAGCCTGGAGGAAATA

GlnLeuValHisSerGluHisHisSerLeuLeuTyrGlyThrAsnProLeuAspGlyGln
2221 CAGCTTGTTCATTCTGAACATCACTCACTGTTGTATGGCACCAACCCCCTGGACGGACAG

LysLeuAspProArgIleLeuLeuGlyAspAspSerGlnLysPhePheSerSerLeuPro
2281 AAGCTGGACCCCAGGATACTCCTAGGTGATGACTCTCAAAAGTTTTTTTCCTCATTACCT

FIG. 22E

```
      CysGlyGlyLeuGlyValAspSerAspThrIleTrpAsnGluLeuHisSerSerGlyAla
2341  TGTGGTGGACTTGGGGTGGACAGTGACACCATTTGGAATGAGCTACACTCGTCCGGTGCT

      AlaArgMetAlaValGlyCysValIleGluLeuAlaSerLysValAlaSerGlyGluLeu
2401  GCACGCATGGCTGTTGGCTGTGTCATCGAGCTGGCTTCCAAAGTGGCCTCAGGAGAGCTG

      LysAsnGlyPheAlaValValArgProProGlyHisHisAlaGluGluSerThrAlaMet
2461  AAGAATGGGTTTGCTGTTGTGAGGCCCCCTGGCCATCACGCTGAAGAATCCACAGCCATG

      GlyPheCysPhePheAsnSerValAlaIleThrAlaLysTyrLeuArgAspGlnLeuAsn
2521  GGGTTCTGCTTTTTTAATTCAGTTGCAATTACCGCCAAATACTTGAGAGACCAACTAAAT

      IleSerLysIleLeuIleValAspLeuAspValHisHisGlyAsnGlyThrGlnGlnAla
2581  ATAAGCAAGATATTGATTGTAGATCTGGATGTTCACCATGGAAACGGTACCCAGCAGGCC

      PheTyrAlaAspProSerIleLeuTyrIleSerLeuHisArgTyrAspGluGlyAsnPhe
2641  TTTTATGCTGACCCCAGCATCCTGTACATTTCACTCCATCGCTATGATGAAGGGAACTTT

      PheProGlySerGlyAlaProAsnGluValArgPheIleSerLeuGluProHisPheTyr
2701  TTCCCTGGCAGTGGAGCCCCAAATGAGGTTCGGTTTATTTCTTTAGAGCCCCACTTTTAT

      LeuTyrLeuSerGlyAsnCysIleAla***
2761  TTGTATCTTTCAGGTAATTGCATTGCATGA
```

FIG. 22F

```
   1  GGGGAAGAGAGGCACAGACACAGATAGGAGAAGGGCACCGGCTGGAGCCACTTGCAGGAC
  61  TGAGGGTTTTTGCAACAAAACCCTAGCAGCCTGAAGAACTCTAAGCCAGATGGGGTGGCT

                                          MetHisSerMetIleSerSerValAspVal
 121  GGACGAGAGCAGCTCTTGGCTCAGCAAAGAATGCACAGTATGATCAGCTCAGTGGATGTG

      LysSerGluValProValGlyLeuGluProIleSerProLeuAspLeuArgThrAspLeu
 181  AAGTCAGAAGTTCCTGTGGGCCTGGAGCCCATCTCACCTTTAGACCTAAGGACAGACCTC

      ArgMetMetMetProValValAspProValValArgGluLysGlnLeuGlnGlnGluLeu
 241  AGGATGATGATGCCCGTGGTGGACCCTGTTGTCCGTGAGAAGCAATTGCAGCAGGAATTA

      LeuLeuIleGlnGlnGlnGlnGlnIleGlnLysGlnLeuLeuIleAlaGluPheGlnLys
 301  CTTCTTATCCAGCAGCAGCAACAAATCCAGAAGCAGCTTCTGATAGCAGAGTTTCAGAAA

      GlnHisGluAsnLeuThrArgGlnHisGlnAlaGlnLeuGlnGluHisIleLysGluLeu
 361  CAGCATGAGAACTTGACACGGCAGCACCAGGCTCAGCTTCAGGAGCATATCAAGGAACTT

      LeuAlaIleLysGlnGlnGlnGluLeuLeuGluLysGluGlnLysLeuGluGlnGlnArg
 421  CTAGCCATAAAACAGCAACAAGAACTCCTAGAAAAGGAGCAGAAACTGGAGCAGCAGAGG

      GlnGluGlnGluValGluArgHisArgArgGluGlnGlnLeuProProLeuArgGlyLys
 481  CAAGAACAGGAAGTAGAGAGGCATCGCAGAGAACAGCAGCTTCCTCCTCTCAGAGGCAAA

      AspArgGlyArgGluArgAlaValAlaSerThrGluValLysGlnLysLeuGlnGluPhe
 541  GATAGAGGACGAGAAAGGGCAGTGGCAAGTACAGAAGTAAAGCAGAAGCTTCAAGAGTTC

      LeuLeuSerLysSerAlaThrLysAspThrProThrAsnGlyLysAsnHisSerValSer
 601  CTACTGAGTAAATCAGCAACGAAAGACACTCCAACTAATGGAAAAAATCATTCCGTGAGC

      ArgHisProLysLeuTrpTyrThrAlaAlaHisHisThrSerLeuAspGlnSerSerPro
 661  CGCCATCCCAAGCTCTGGTACACGGCTGCCCACCACACATCATTGGATCAAAGCTCTCCA

      ProLeuSerGlyThrSerProSerTyrLysTyrThrLeuProGlyAlaGlnAspAlaLys
 721  CCCCTTAGTGGAACATCTCCATCCTACAAGTACACATTACCAGGAGCACAAGATGCAAAG

      AspAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArgLeu
 781  GATGATTTCCCCCTTCGAAAAACTGCCTCTGAGCCCAACTTGAAGGTGCGGTCCAGGTTA

      LysGlnLysValAlaGluArgArgSerSerProLeuLeuArgArgLysAspGlyAsnVal
 841  AAACAGAAAGTGGCAGAGAGGAGAAGCAGCCCCTTACTCAGGCGGAAGGATGGAAATGTT

      ValThrSerPheLysLysArgMetPheGluValThrGluSerSerValSerSerSerSer
 901  GTCACTTCATTCAAGAAGCGAATGTTTGAGGTGACAGAATCCTCAGTCAGTAGCAGTTCT

      ProGlySerGlyProSerSerProAsnAsnGlyProThrGlySerValThrGluAsnGlu
 961  CCAGGCTCTGGTCCCAGTTCACCAAACAATGGGCCAACTGGAAGTGTTACTGAAAATGAG

      ThrSerValLeuProProThrProHisAlaGluGlnMetValSerGlnGlnArgIleLeu
1021  ACTTCGGTTTTGCCCCCTACCCCTCATGCCGAGCAAATGGTTTCACAGCAACGCATTCTA

      IleHisGluAspSerMetAsnLeuLeuSerLeuTyrThrSerProSerLeuProAsnIle
1081  ATTCATGAAGATTCCATGAACCTGCTAAGTCTTTATACCTCTCCTTCTTTGCCCAACATT

      ThrLeuGlyLeuProAlaValProSerGlnLeuAsnAlaSerAsnSerLeuLysGluLys
1141  ACCTTGGGGCTTCCCGCAGTGCCATCCCAGCTCAATGCTTCGAATTCACTCAAAGAAAAG
```

FIG. 22G

GlnLysCysGluThrGlnThrLeuArgGlnGlyValProLeuProGlyGlnTyrGlyGly
1201 CAGAAGTGTGAGACGCAGACGCTTAGGCAAGGTGTTCCTCTGCCTGGGCAGTATGGAGGC

SerIleProAlaSerSerSerHisProHisValThrLeuGluGlyLysProProAsnSer
1261 AGCATCCCGGCATCTTCCAGCCACCCTCATGTTACTTTAGAGGGAAAGCCACCCAACAGC

SerHisGlnAlaLeuLeuGlnHisLeuLeuLeuLysGluGlnMetArgGlnGlnLysLeu
1321 AGCCACCAGGCTCTCCTGCAGCATTTATTATTGAAAGAACAAATGCGACAGCAAAAGCTT

LeuValAlaGlyGlyValProLeuHisProGlnSerProLeuAlaThrLysGluArgIle
1381 CTTGTAGCTGGTGGAGTTCCCTTACATCCTCAGTCTCCCTTGGCAACAAAAGAGAGAATT

SerProGlyIleArgGlyThrHisLysLeuProArgHisArgProLeuAsnArgThrGln
1441 TCACCTGGCATTAGAGGTACCCACAAATTGCCCCGTCACAGACCCCTGAACCGAACCCAG

SerAlaProLeuProGlnSerThrLeuAlaGlnLeuValIleGlnGlnGlnHisGlnGln
1501 TCTGCACCTTTGCCTCAGAGCACGTTGGCTCAGCTGGTCATTCAACAGCAACACCAGCAA

PheLeuGluLysGlnLysGlnTyrGlnGlnGlnIleHisMetAsnLysLeuLeuSerLys
1561 TTCTTGGAGAAGCAGAAGCAATACCAGCAGCAGATCCACATGAACAAACTGCTTTCGAAA

SerIleGluGlnLeuLysGlnProGlySerHisLeuGluGluAlaGluGluGluLeuGln
1621 TCTATTGAACAACTGAAGCAACCAGGCAGTCACCTTGAGGAAGCAGAGGAAGAGCTTCAG

GlyAspGlnAlaMetGlnGluAspArgAlaProSerSerGlyAsnSerThrArgSerAsp
1681 GGGGACCAGGCGATGCAGGAAGACAGAGCGCCCTCTAGTGGCAACAGCACTAGGAGCGAC

SerSerAlaCysValAspAspThrLeuGlyGlnValGlyAlaValLysValLysGluGlu
1741 AGCAGTGCTTGTGTGGATGACACACTGGGACAAGTTGGGGCTGTGAAGGTCAAGGAGGAA

ProValAspSerAspGluAspAlaGlnIleGlnGluMetGluSerGlyGluGlnAlaAla
1801 CCAGTGGACAGTGATGAAGATGCTCAGATCCAGGAAATGGAATCTGGGGAGCAGGCTGCT

PheMetGlnGlnValIleGlyLysAspLeuAlaProGlyPheValIleLysValIleIle
1861 TTTATGCAACAGGTAATAGGCAAAGATTTAGCTCCAGGATTTGTAATTAAAGTCATTATC

***
1921 TGAACATGAAATGCATTGCAGGTTTGGTAAATGGATATGATTTCCTATCAGTTTATATTT

1981 CTCTATGATTTGAGTTCAGTGTTTAAGGATTCTACCTAATGCAGATATATGTATATATCT
2041 ATATAGAGGTCTTTCTATATACTGATCTCTATATAGATATCAATGTTTCATTGAAAATCC
2101 ACTGGTAAGGAAATACCTGTTATACTAAAATTATGATACATAATATCTGAGCAGTTAATA
2161 GGCTTTAAATTTATCCCAAAGCCTGCTACACCAATTACTTCTAAAGAAAACAAATTCACT
2221 GTTATTTTGAGTTTATGTGTTGAGATCAGTGACTGCTGGATAGTCTCCCAGTCTGATCAA
2281 TGAAGCATTCGATTAGTTTTTGATTTTTTGCAACATCTAGAATTTAATTTTCACATCACT
2341 GTACATAATGTATCATACTATAGTCTTGAACACTGTTAAAGGTAGTCTGCCCCTTCCTTC
2401 CTCTCTCTTTTTTTAGTTAAGTAGAAATGTTCTGGTCACCATGCCAGTAGTCCTAGGTTA
2461 TTGTGTAGGTTGCAATTGAACATATTAGGAATACAGGTGGTTTTAAATATATAGATGCAA
2521 ATTGCAGCACTACTTTAAATATTAGATTATGTCTCACATAGCACTGCTCATTTTACTTTT
2581 ATTTTGTGTAATTTGATGACACTGTCTATCAAAAAAGAGCAAATGAAGCAGATGCAAATG
2641 TTAGTGAGAAGTAATGTGCAGCATTATGGTCCAATCAGATACAATATTGTGTCTACAATT
2701 GCAAAAAACACAGTAACAGGATGAATATTATCTGATATCAAGTCAAAATCAGTTTGAAAA
2761 GAAGGTGTATCATATTTTATATTGTCACTAGAATCTCTTAAGTATAATTCCATAATGACA
2821 TGGGCATATACCGTAACATTCTGGCAAATAACAATTAGAAAAGATAGGTTTAACAAAAAA
2881 ATTTACTTGTATATAATGCACCTTCAGGAGGACTATGTCCTTTGATGCTATAAAATACAA
2941 ACAACTTTGAAGGCAACAGAAGACACTGTTTATTCAAGTCAGTTCTTTGTCAGGTTCCTG
3001 CTGTTCTCCTACAGAAAAGTGATTCTGTGAGGGTGAACAGGAAATGCCTTGTGGAAACAG
3061 GAAGTCCAAGTGATTCATGTACTGAGGAATGTAGGAAAAAAAATCTGAGGATAGTGCTTT

FIG. 22H

3121 ACTCTTTCTGTTTTTAAAGGGCACTCTATGAATTGATTTATTGTCTAAGAAAATAACACC
3181 ACAAGTAGGGAAATTGTTACGGAAGCTTTTCACTGGAACATTTCCTTCATATTCCCTTTT
3241 GATATGTTTACCTTGTTTTATAGGTTTACTTTTGTTAAGCTAGTTAAAGGTTCGTTGTAT
3301 TAAGACCCCTTTAATATGGATAATCCAAATTGACCTAGAATCTTTGTGAGGTTTTTTCTA
3361 TTAAAATATTTATATTTCTAAATCCGAGGTATTTCAAGGTGTAGTATCCTATTTCAAAGG
3421 AGATATAGCAGTTTTGCCAAATGTAGACATTGTTCAACTGTATGTTATTGGCACGTGTTG
3481 TTTACATTTTGCTGTGACATTTAAAAATATTTCTTTAAAAATGTTACTGCTAAAGATACA
3541 TTATCCTTTTTTAAAAAGTCTCCATTCAAATTAAATTAACATAACTAGAAGTTAGAAAGT
3601 TTAAAAGTTTTCCACATAATGAAAGTCCTTCTGATAATTTGACAAATAGCTATAATAGGA
3661 ACACTCCCTATCACCAACATATTTTGGTTAGTATATTCCTTCATATTAAAATGACTTTTT
3721 GTCAGTTGTTTTGCATTAAAAATATGGCATGCCTAAGATAAAATTGTATATTTTTTCCAT
3781 CTCATAAATATTCATTTTCTTCAAAGTCTTTTTTCAATCTCATAAAAAAGGGATAGTGCA
3841 TCTTTTAAAATACATTTTATTTGGGGAGGAACATGTGGCTGAGCAGACTTTTGTATAATA
3901 TTACTTCAAAGATATGTAATCACAAACAAAAAAAACTATTTTTTATAATGTCATTTGAGA
3961 GAGTTTCATCAGTACAGTTGGTGGACGTTAATTGTTTGAATTTGATAGTCTTTGAATTTA
4021 ATCAAGAAACTACCTGGAACCAGTGAAAAGGAAAGCTGGACTTAAATAATCTTAGAATTA
4081 ATTGATAAATGTCTCTTTTAAAATCTACTGTATTTATTATAATTTACACCCTTGAAGGTG
4141 ATCTCTTGTTTTGTGTTGTAAATATATTGTTTGTATGTTTCCCTTCTTGCCTTCTGTTAT
4201 AAGTCTCTTCCTTTCTCAAATAAAGTTTTTTTTAAAAG

FIG. 22I

```
                            1                                                50
BMY_HDACX_V1        (1)   --------------------------------------------------
BMY_HDACX_V2        (1)   CCACGCGTCCGTAGGAGAAGGGCACCGGCTGGAGCCACTTGCAGGACTGA
HDAC9V1             (1)   --------------------------------------------------
HDAC9V2             (1)   --------------------------------------------------
HDAC9V3             (1)   --------------------------------------------------
CONSENSUS           (1)
                            51                                              100
BMY_HDACX_V1        (1)   --------------------------------------------------
BMY_HDACX_V2       (51)   GGGTTTTTGCAACAAAACCCTAGCAGCCTGAAGAACTCTAAGCCAGGTTT
HDAC9V1             (1)   --------------------------------------------------
HDAC9V2             (1)   --------------------------------------------------
HDAC9V3             (1)   --------------------------------------------------
CONSENSUS          (51)
                            101                                             150
BMY_HDACX_V1        (1)   --------------------------------------------------
BMY_HDACX_V2      (101)   AATTGGTTTCTTTTTCTCGTGGGTAGACTTAATAATTTTCTACGTATTCT
HDAC9V1             (1)   --------------------------------------------------
HDAC9V2             (1)   --------------------------------------------------
HDAC9V3             (1)   --------------------------------------------------
CONSENSUS         (101)
                            151                                             200
BMY_HDACX_V1        (1)   --------------------------------------------------
BMY_HDACX_V2      (151)   GACAAAGAAATAACCCCGAAGCACGTTCCTATTTCCCACCTGCTTGTAGT
HDAC9V1             (1)   ------GGGGAAGAGAGGCACAGACACAGATAGGAGAAGGGCACCGGCTG
HDAC9V2             (1)   ------GGGGAAGAGAGGCACAGACACAGATAGGAGAAGGGCACCGGCTG
HDAC9V3             (1)   ------GGGGAAGAGAGGCACAGACACAGATAGGAGAAGGGCACCGGCTG
CONSENSUS         (151)         GGGGAAGAGAGGCACAGACACAGATAGGAGAAGGGCACCGGCTG
                            201                                             250
BMY_HDACX_V1        (1)   --------------------------------------------------
BMY_HDACX_V2      (201)   TTCCGGGATAACCTAAACTCCAGAGAGCTATAGCATCCACTCTGTCCTTT
HDAC9V1            (45)   GAGCCACTTGCAGGACTGAGGGTTTTTGCAACAAAACCCTAGCAGCCTGA
HDAC9V2            (45)   GAGCCACTTGCAGGACTGAGGGTTTTTGCAACAAAACCCTAGCAGCCTGA
HDAC9V3            (45)   GAGCCACTTGCAGGACTGAGGGTTTTTGCAACAAAACCCTAGCAGCCTGA
CONSENSUS         (201)   GAGCCACTTGCAGGACTGAGGGTTTTTGCAACAAAACCCTAGCAGCCTGA
                            251                                             300
BMY_HDACX_V1        (1)   --------------------------------------------------
BMY_HDACX_V2      (251)   CTGCTTTGCACACAGATGGGGTGGCTGGACGAGAGCAGCTCTTGGCTCAG
HDAC9V1            (95)   AGAACTCTAAGCCAGATGGGGTGGCTGGACGAGAGCAGCTCTTGGCTCAG
HDAC9V2            (95)   AGAACTCTAAGCCAGATGGGGTGGCTGGACGAGAGCAGCTCTTGGCTCAG
HDAC9V3            (95)   AGAACTCTAAGCCAGATGGGGTGGCTGGACGAGAGCAGCTCTTGGCTCAG
CONSENSUS         (251)   AGAACTCTAAGCCAGATGGGGTGGCTGGACGAGAGCAGCTCTTGGCTCAG
                                           * SPLICE JUNCTION: CAG>>>ATG
                            301                                             350
BMY_HDACX_V1        (1)   --------------------------------------------------
BMY_HDACX_V2      (301)   CAAAGAATGCACAGTATGATCAGCTCAGTGGATGTGAAGTCAGAAGTTCC
HDAC9V1           (145)   CAAAGAATGCACAGTATGATCAGCTCAGTGGATGTGAAGTCAGAAGTTCC
HDAC9V2           (145)   CAAAGAATGCACAGTATGATCAGCTCAGTGGATGTGAAGTCAGAAGTTCC
HDAC9V3           (145)   CAAAGAATGCACAGTATGATCAGCTCAGTGGATGTGAAGTCAGAAGTTCC
CONSENSUS         (301)   CAAAGAATGCACAGTATGATCAGCTCAGTGGATGTGAAGTCAGAAGTTCC
                            351                                             400
BMY_HDACX_V1        (1)   --------------------------------------------------
BMY_HDACX_V2      (351)   TGTGGGCCTGGAGCCCATCTCACCTTTAGACCTAAGGACAGACCTCAGGA
HDAC9V1           (195)   TGTGGGCCTGGAGCCCATCTCACCTTTAGACCTAAGGACAGACCTCAGGA
HDAC9V2           (195)   TGTGGGCCTGGAGCCCATCTCACCTTTAGACCTAAGGACAGACCTCAGGA
HDAC9V3           (195)   TGTGGGCCTGGAGCCCATCTCACCTTTAGACCTAAGGACAGACCTCAGGA
CONSENSUS         (351)   TGTGGGCCTGGAGCCCATCTCACCTTTAGACCTAAGGACAGACCTCAGGA
```

FIG. 23A

```
                              401                                              450
BMY_HDACX_V1      (1)   --------------------------------------------------
BMY_HDACX_V2    (401)   TGATGATGCCCGTGGTGGACCCTGTTGTCCGTGAGAAGCAATTGCAGCAG
HDAC9V1         (245)   TGATGATGCCCGTGGTGGACCCTGTTGTCCGTGAGAAGCAATTGCAGCAG
HDAC9V2         (245)   TGATGATGCCCGTGGTGGACCCTGTTGTCCGTGAGAAGCAATTGCAGCAG
HDAC9V3         (245)   TGATGATGCCCGTGGTGGACCCTGTTGTCCGTGAGAAGCAATTGCAGCAG
CONSENSUS       (401)   TGATGATGCCCGTGGTGGACCCTGTTGTCCGTGAGAAGCAATTGCAGCAG
                        451                                              500
BMY_HDACX_V1      (1)   --------------------------------------------------
BMY_HDACX_V2    (451)   GAATTACTTCTTATCCAGCAGCAGCAACAAATCCAGAAGCAGCTTCTGAT
HDAC9V1         (295)   GAATTACTTCTTATCCAGCAGCAGCAACAAATCCAGAAGCAGCTTCTGAT
HDAC9V2         (295)   GAATTACTTCTTATCCAGCAGCAGCAACAAATCCAGAAGCAGCTTCTGAT
HDAC9V3         (295)   GAATTACTTCTTATCCAGCAGCAGCAACAAATCCAGAAGCAGCTTCTGAT
CONSENSUS       (451)   GAATTACTTCTTATCCAGCAGCAGCAACAAATCCAGAAGCAGCTTCTGAT
                        501                                              550
BMY_HDACX_V1      (1)   --------------------------------------------------
BMY_HDACX_V2    (501)   AGCAGAGTTTCAGAAACAGCATGAGAACTTGACACGGCAGCACCAGGCTC
HDAC9V1         (345)   AGCAGAGTTTCAGAAACAGCATGAGAACTTGACACGGCAGCACCAGGCTC
HDAC9V2         (345)   AGCAGAGTTTCAGAAACAGCATGAGAACTTGACACGGCAGCACCAGGCTC
HDAC9V3         (345)   AGCAGAGTTTCAGAAACAGCATGAGAACTTGACACGGCAGCACCAGGCTC
CONSENSUS       (501)   AGCAGAGTTTCAGAAACAGCATGAGAACTTGACACGGCAGCACCAGGCTC
                        551                                              600
BMY_HDACX_V1      (1)   --------------------------------------------------
BMY_HDACX_V2    (551)   AGCTTCAGGAGCATATCAAGTTGCAACAGGAACTTCTAGCCATAAAACAG
HDAC9V1         (395)   AGCTTCAGGAGCATATCAAG---------GAACTTCTAGCCATAAAACAG
HDAC9V2         (395)   AGCTTCAGGAGCATATCAAG---------GAACTTCTAGCCATAAAACAG
HDAC9V3         (395)   AGCTTCAGGAGCATATCAAG---------GAACTTCTAGCCATAAAACAG
CONSENSUS       (551)   AGCTTCAGGAGCATATCAAG         GAACTTCTAGCCATAAAACAG
                                            *SPLICE ACCEPTOR I
                                                     *SPLICE ACCEPTOR 2
                        601                                              650
BMY_HDACX_V1      (1)   --------------------------------------------------
BMY_HDACX_V2    (601)   CAACAAGAACTCCTAGAAAAGGAGCAGAAACTGGAGCAGCAGAGGCAAGA
HDAC9V1         (436)   CAACAAGAACTCCTAGAAAAGGAGCAGAAACTGGAGCAGCAGAGGCAAGA
HDAC9V2         (436)   CAACAAGAACTCCTAGAAAAGGAGCAGAAACTGGAGCAGCAGAGGCAAGA
HDAC9V3         (436)   CAACAAGAACTCCTAGAAAAGGAGCAGAAACTGGAGCAGCAGAGGCAAGA
CONSENSUS       (601)   CAACAAGAACTCCTAGAAAAGGAGCAGAAACTGGAGCAGCAGAGGCAAGA
                        651                                              700
BMY_HDACX_V1      (1)   --------------------------------------------------
BMY_HDACX_V2    (651)   ACAGGAAGTAGAGAGGCATCGCAGAGAACAGCAGCTTCCTCCTCTCAGAG
HDAC9V1         (486)   ACAGGAAGTAGAGAGGCATCGCAGAGAACAGCAGCTTCCTCCTCTCAGAG
HDAC9V2         (486)   ACAGGAAGTAGAGAGGCATCGCAGAGAACAGCAGCTTCCTCCTCTCAGAG
HDAC9V3         (486)   ACAGGAAGTAGAGAGGCATCGCAGAGAACAGCAGCTTCCTCCTCTCAGAG
CONSENSUS       (651)   ACAGGAAGTAGAGAGGCATCGCAGAGAACAGCAGCTTCCTCCTCTCAGAG
                        701                                              750
BMY_HDACX_V1      (1)   --------------------------------------------------
BMY_HDACX_V2    (701)   GCAAAGATAGAGGACGAGAAAGGGCAGTGGCAAGTACAGAAGTAAAGCAG
HDAC9V1         (536)   GCAAAGATAGAGGACGAGAAAGGGCAGTGGCAAGTACAGAAGTAAAGCAG
HDAC9V2         (536)   GCAAAGATAGAGGACGAGAAAGGGCAGTGGCAAGTACAGAAGTAAAGCAG
HDAC9V3         (536)   GCAAAGATAGAGGACGAGAAAGGGCAGTGGCAAGTACAGAAGTAAAGCAG
CONSENSUS       (701)   GCAAAGATAGAGGACGAGAAAGGGCAGTGGCAAGTACAGAAGTAAAGCAG
                        751                                              800
BMY_HDACX_V1      (1)   --------------------------------------------------
BMY_HDACX_V2    (751)   AAGCTTCAAGAGTTCCTACTGAGTAAATCAGCAACGAAAGACACTCCAAC
HDAC9V1         (586)   AAGCTTCAAGAGTTCCTACTGAGTAAATCAGCAACGAAAGACACTCCAAC
HDAC9V2         (586)   AAGCTTCAAGAGTTCCTACTGAGTAAATCAGCAACGAAAGACACTCCAAC
HDAC9V3         (586)   AAGCTTCAAGAGTTCCTACTGAGTAAATCAGCAACGAAAGACACTCCAAC
CONSENSUS       (751)   AAGCTTCAAGAGTTCCTACTGAGTAAATCAGCAACGAAAGACACTCCAAC
                        801                                              850
BMY_HDACX_V1      (1)   --------------------------------------------------
BMY_HDACX_V2    (801)   TAATGGAAAAAATCATTCCGTGAGCCGCCATCCCAAGCTCTGGTACACGG
HDAC9V1         (636)   TAATGGAAAAAATCATTCCGTGAGCCGCCATCCCAAGCTCTGGTACACGG
HDAC9V2         (636)   TAATGGAAAAAATCATTCCGTGAGCCGCCATCCCAAGCTCTGGTACACGG
HDAC9V3         (636)   TAATGGAAAAAATCATTCCGTGAGCCGCCATCCCAAGCTCTGGTACACGG
CONSENSUS       (801)   TAATGGAAAAAATCATTCCGTGAGCCGCCATCCCAAGCTCTGGTACACGG
```

FIG. 23B

```
                       851                                              900
BMY_HDACX_V1     (1)   --------------------------------------------------
BMY_HDACX_V2   (851)   CTGCCCACCACACATCATTGGATCAAAGCTCTCCACCCCTTAGTGGAACA
     HDAC9V1   (686)   CTGCCCACCACACATCATTGGATCAAAGCTCTCCACCCCTTAGTGGAACA
     HDAC9V2   (686)   CTGCCCACCACACATCATTGGATCAAAGCTCTCCACCCCTTAGTGGAACA
     HDAC9V3   (686)   CTGCCCACCACACATCATTGGATCAAAGCTCTCCACCCCTTAGTGGAACA
   CONSENSUS   (851)   CTGCCCACCACACATCATTGGATCAAAGCTCTCCACCCCTTAGTGGAACA
                       901                                              950
BMY_HDACX_V1     (1)   --------------------------------------------------
BMY_HDACX_V2   (901)   TCTCCATCCTACAAGTACACATTACCAGGAGCACAAGATGCAAAGGATGA
     HDAC9V1   (736)   TCTCCATCCTACAAGTACACATTACCAGGAGCACAAGATGCAAAGGATGA
     HDAC9V2   (736)   TCTCCATCCTACAAGTACACATTACCAGGAGCACAAGATGCAAAGGATGA
     HDAC9V3   (736)   TCTCCATCCTACAAGTACACATTACCAGGAGCACAAGATGCAAAGGATGA
   CONSENSUS   (901)   TCTCCATCCTACAAGTACACATTACCAGGAGCACAAGATGCAAAGGATGA
                       951                                             1000
BMY_HDACX_V1     (1)   --------------------------------------------------
BMY_HDACX_V2   (951)   TTTCCCCCTTCGAAAAACTGCCTCTGAGCCCAACTTGAAGGTGCGGTCCA
     HDAC9V1   (786)   TTTCCCCCTTCGAAAAACTGCCTCTGAGCCCAACTTGAAGGTGCGGTCCA
     HDAC9V2   (786)   TTTCCCCCTTCGAAAAACTGCCTCTGAGCCCAACTTGAAGGTGCGGTCCA
     HDAC9V3   (786)   TTTCCCCCTTCGAAAAACTGCCTCTGAGCCCAACTTGAAGGTGCGGTCCA
   CONSENSUS   (951)   TTTCCCCCTTCGAAAAACTGCCTCTGAGCCCAACTTGAAGGTGCGGTCCA
                       1001                                            1050
BMY_HDACX_V1     (1)   --------------------------------------------------
BMY_HDACX_V2  (1001)   GGTTAAAACAGAAAGTGGCAGAGAGGAGAAGCAGCCCCTTACTCAGGCGG
     HDAC9V1   (836)   GGTTAAAACAGAAAGTGGCAGAGAGGAGAAGCAGCCCCTTACTCAGGCGG
     HDAC9V2   (836)   GGTTAAAACAGAAAGTGGCAGAGAGGAGAAGCAGCCCCTTACTCAGGCGG
     HDAC9V3   (836)   GGTTAAAACAGAAAGTGGCAGAGAGGAGAAGCAGCCCCTTACTCAGGCGG
   CONSENSUS  (1001)   GGTTAAAACAGAAAGTGGCAGAGAGGAGAAGCAGCCCCTTACTCAGGCGG
                       1051                                            1100
BMY_HDACX_V1     (1)   --------------------------------------------------
BMY_HDACX_V2  (1051)   AAGGATGGAAATGTTGTCACTTCATTCAAGAAGCGAATGTTTGAGGTGAC
     HDAC9V1   (886)   AAGGATGGAAATGTTGTCACTTCATTCAAGAAGCGAATGTTTGAGGTGAC
     HDAC9V2   (886)   AAGGATGGAAATGTTGTCACTTCATTCAAGAAGCGAATGTTTGAGGTGAC
     HDAC9V3   (886)   AAGGATGGAAATGTTGTCACTTCATTCAAGAAGCGAATGTTTGAGGTGAC
   CONSENSUS  (1051)   AAGGATGGAAATGTTGTCACTTCATTCAAGAAGCGAATGTTTGAGGTGAC
                       1101                                            1150
BMY_HDACX_V1     (1)   --------------------------------------------------
BMY_HDACX_V2  (1101)   AGAATCCTCAGTCAGTAGCAGTTCTCCAGGCTCTGGTCCCAGTTCACCAA
     HDAC9V1   (936)   AGAATCCTCAGTCAGTAGCAGTTCTCCAGGCTCTGGTCCCAGTTCACCAA
     HDAC9V2   (936)   AGAATCCTCAGTCAGTAGCAGTTCTCCAGGCTCTGGTCCCAGTTCACCAA
     HDAC9V3   (936)   AGAATCCTCAGTCAGTAGCAGTTCTCCAGGCTCTGGTCCCAGTTCACCAA
   CONSENSUS  (1101)   AGAATCCTCAGTCAGTAGCAGTTCTCCAGGCTCTGGTCCCAGTTCACCAA
                       1151                                            1200
BMY_HDACX_V1     (1)   -----------------------GCTGAAAATGAGACTTCGGTTTTGCCC
BMY_HDACX_V2  (1151)   ACAATGGGCCAACTGGAAGTGTTACTGAAAATGAGACTTCGGTTTTGCCC
     HDAC9V1   (986)   ACAATGGGCCAACTGGAAGTGTTACTGAAAATGAGACTTCGGTTTTGCCC
     HDAC9V2   (986)   ACAATGGGCCAACTGGAAGTGTTACTGAAAATGAGACTTCGGTTTTGCCC
     HDAC9V3   (986)   ACAATGGGCCAACTGGAAGTGTTACTGAAAATGAGACTTCGGTTTTGCCC
   CONSENSUS  (1151)   ACAATGGGCCAACTGGAAGTGTTACTGAAAATGAGACTTCGGTTTTGCCC
                       1201                                            1250
BMY_HDACX_V1    (28)   CCTACCCCTCATGCCGAGCAAATGGTTTCACAGCAACGCATTCTAATTCA
BMY_HDACX_V2  (1201)   CCTACCCCTCATGCCGAGCAAATGGTTTCACAGCAACGCATTCTAATTCA
     HDAC9V1  (1036)   CCTACCCCTCATGCCGAGCAAATGGTTTCACAGCAACGCATTCTAATTCA
     HDAC9V2  (1036)   CCTACCCCTCATGCCGAGCAAATGGTTTCACAGCAACGCATTCTAATTCA
     HDAC9V3  (1036)   CCTACCCCTCATGCCGAGCAAATGGTTTCACAGCAACGCATTCTAATTCA
   CONSENSUS  (1201)   CCTACCCCTCATGCCGAGCAAATGGTTTCACAGCAACGCATTCTAATTCA
                       1251                                            1300
BMY_HDACX_V1    (78)   TGAAGATTCCATGAACCTGCTAAGTCTTTATACCTCTCCTTCTTTGCCCA
BMY_HDACX_V2  (1251)   TGAAGATTCCATGAACCTGCTAAGTCTTTATACCTCTCCTTCTTTGCCCA
     HDAC9V1  (1086)   TGAAGATTCCATGAACCTGCTAAGTCTTTATACCTCTCCTTCTTTGCCCA
     HDAC9V2  (1086)   TGAAGATTCCATGAACCTGCTAAGTCTTTATACCTCTCCTTCTTTGCCCA
     HDAC9V3  (1086)   TGAAGATTCCATGAACCTGCTAAGTCTTTATACCTCTCCTTCTTTGCCCA
   CONSENSUS  (1251)   TGAAGATTCCATGAACCTGCTAAGTCTTTATACCTCTCCTTCTTTGCCCA
```

FIG. 23C

```
                          1301                                              1350
BMY_HDACX_V1      (128)   ACATTACCTTGGGGCTTCCCGCAGTGCCATCCCAGCTCAATGCTTCGAAT
BMY_HDACX_V2     (1301)   ACATTACCTTGGGGCTTCCCGCAGTGCCATCCCAGCTCAATGCTTCGAAT
HDAC9V1          (1136)   ACATTACCTTGGGGCTTCCCGCAGTGCCATCCCAGCTCAATGCTTCGAAT
HDAC9V2          (1136)   ACATTACCTTGGGGCTTCCCGCAGTGCCATCCCAGCTCAATGCTTCGAAT
HDAC9V3          (1136)   ACATTACCTTGGGGCTTCCCGCAGTGCCATCCCAGCTCAATGCTTCGAAT
CONSENSUS        (1301)   ACATTACCTTGGGGCTTCCCGCAGTGCCATCCCAGCTCAATGCTTCGAAT
                          1351                                              1400
BMY_HDACX_V1      (178)   TCACTCAAAGAAAAGCAGAAGTGTGAGACGCAGACGCTTAGGCAAGGTGT
BMY_HDACX_V2     (1351)   TCACTCAAAGAAAAGCAGAAGTGTGAGACGCAGACGCTTAGGCAAGGTGT
HDAC9V1          (1186)   TCACTCAAAGAAAAGCAGAAGTGTGAGACGCAGACGCTTAGGCAAGGTGT
HDAC9V2          (1186)   TCACTCAAAGAAAAGCAGAAGTGTGAGACGCAGACGCTTAGGCAAGGTGT
HDAC9V3          (1186)   TCACTCAAAGAAAAGCAGAAGTGTGAGACGCAGACGCTTAGGCAAGGTGT
CONSENSUS        (1351)   TCACTCAAAGAAAAGCAGAAGTGTGAGACGCAGACGCTTAGGCAAGGTGT
                          1401                                              1450
BMY_HDACX_V1      (228)   TCCTCTGCCTGGGCAGTATGGAGGCAGCATCCCGGCATCTTCCAGCCACC
BMY_HDACX_V2     (1401)   TCCTCTGCCTGGGCAGTATGGAGGCAGCATCCCGGCATCTTCCAGCCACC
HDAC9V1          (1236)   TCCTCTGCCTGGGCAGTATGGAGGCAGCATCCCGGCATCTTCCAGCCACC
HDAC9V2          (1236)   TCCTCTGCCTGGGCAGTATGGAGGCAGCATCCCGGCATCTTCCAGCCACC
HDAC9V3          (1236)   TCCTCTGCCTGGGCAGTATGGAGGCAGCATCCCGGCATCTTCCAGCCACC
CONSENSUS        (1401)   TCCTCTGCCTGGGCAGTATGGAGGCAGCATCCCGGCATCTTCCAGCCACC
                          1451                                              1500
BMY_HDACX_V1      (278)   CTCATGTTACTTTAGAGGGAAAGCCACCCAACAGCAGCCACCAGGCTCTC
BMY_HDACX_V2     (1451)   CTCATGTTACTTTAGAGGGAAAGCCACCCAACAGCAGCCACCAGGCTCTC
HDAC9V1          (1286)   CTCATGTTACTTTAGAGGGAAAGCCACCCAACAGCAGCCACCAGGCTCTC
HDAC9V2          (1286)   CTCATGTTACTTTAGAGGGAAAGCCACCCAACAGCAGCCACCAGGCTCTC
HDAC9V3          (1286)   CTCATGTTACTTTAGAGGGAAAGCCACCCAACAGCAGCCACCAGGCTCTC
CONSENSUS        (1451)   CTCATGTTACTTTAGAGGGAAAGCCACCCAACAGCAGCCACCAGGCTCTC
                          1501                                              1550
BMY_HDACX_V1      (328)   CTGCAGCATTTATTATTGAAAGAACAAATGCGACAGCAAAAGCTTCTTGT
BMY_HDACX_V2     (1501)   CTGCAGCATTTATTATTGAAAGAACAAATGCGACAGCAAAAGCTTCTTGT
HDAC9V1          (1336)   CTGCAGCATTTATTATTGAAAGAACAAATGCGACAGCAAAAGCTTCTTGT
HDAC9V2          (1336)   CTGCAGCATTTATTATTGAAAGAACAAATGCGACAGCAAAAGCTTCTTGT
HDAC9V3          (1336)   CTGCAGCATTTATTATTGAAAGAACAAATGCGACAGCAAAAGCTTCTTGT
CONSENSUS        (1501)   CTGCAGCATTTATTATTGAAAGAACAAATGCGACAGCAAAAGCTTCTTGT
                          1551                                              1600
BMY_HDACX_V1      (378)   AGCTGGTGGAGTTCCCTTACATCCTCAGTCTCCCTTGGCAACAAAAGAGA
BMY_HDACX_V2     (1551)   AGCTGGTGGAGTTCCCTTACATCCTCAGTCTCCCTTGGCAACAAAAGAGA
HDAC9V1          (1386)   AGCTGGTGGAGTTCCCTTACATCCTCAGTCTCCCTTGGCAACAAAAGAGA
HDAC9V2          (1386)   AGCTGGTGGAGTTCCCTTACATCCTCAGTCTCCCTTGGCAACAAAAGAGA
HDAC9V3          (1386)   AGCTGGTGGAGTTCCCTTACATCCTCAGTCTCCCTTGGCAACAAAAGAGA
CONSENSUS        (1551)   AGCTGGTGGAGTTCCCTTACATCCTCAGTCTCCCTTGGCAACAAAAGAGA
                          1601                                              1650
BMY_HDACX_V1      (428)   GAATTTCACCTGGCATTAGAGGTACCCACAAATTGCCCCGTCACAGACCC
BMY_HDACX_V2     (1601)   GAATTTCACCTGGCATTAGAGGTACCCACAAATTGCCCCGTCACAGACCC
HDAC9V1          (1436)   GAATTTCACCTGGCATTAGAGGTACCCACAAATTGCCCCGTCACAGACCC
HDAC9V2          (1436)   GAATTTCACCTGGCATTAGAGGTACCCACAAATTGCCCCGTCACAGACCC
HDAC9V3          (1436)   GAATTTCACCTGGCATTAGAGGTACCCACAAATTGCCCCGTCACAGACCC
CONSENSUS        (1601)   GAATTTCACCTGGCATTAGAGGTACCCACAAATTGCCCCGTCACAGACCC
                          1651                                              1700
BMY_HDACX_V1      (478)   CTGAACCGAACCCAGTCTGCACCTTTGCCTCAGAGCACGTTGGCTCAGCT
BMY_HDACX_V2     (1651)   CTGAACCGAACCCAGTCTGCACCTTTGCCTCAGAGCACGTTGGCTCAGCT
HDAC9V1          (1486)   CTGAACCGAACCCAGTCTGCACCTTTGCCTCAGAGCACGTTGGCTCAGCT
HDAC9V2          (1486)   CTGAACCGAACCCAGTCTGCACCTTTGCCTCAGAGCACGTTGGCTCAGCT
HDAC9V3          (1486)   CTGAACCGAACCCAGTCTGCACCTTTGCCTCAGAGCACGTTGGCTCAGCT
CONSENSUS        (1651)   CTGAACCGAACCCAGTCTGCACCTTTGCCTCAGAGCACGTTGGCTCAGCT
                          1701                                              1750
BMY_HDACX_V1      (528)   GGTCATTCAACAGCAACACCAGCAATTCTTGGAGAAGCAGAAGCAATACC
BMY_HDACX_V2     (1701)   GGTCATTCAACAGCAACACCAGCAATTCTTGGAGAAGCAGAAGCAATACC
HDAC9V1          (1536)   GGTCATTCAACAGCAACACCAGCAATTCTTGGAGAAGCAGAAGCAATACC
HDAC9V2          (1536)   GGTCATTCAACAGCAACACCAGCAATTCTTGGAGAAGCAGAAGCAATACC
HDAC9V3          (1536)   GGTCATTCAACAGCAACACCAGCAATTCTTGGAGAAGCAGAAGCAATACC
CONSENSUS        (1701)   GGTCATTCAACAGCAACACCAGCAATTCTTGGAGAAGCAGAAGCAATACC
```

FIG. 23D

```
                        1751                                              1800
BMY_HDACX_V1    (578)   AGCAGCAGATCCACATGAACAAACTGCTTTCGAAATCTATTGAACAACTG
BMY_HDACX_V2   (1751)   AGCAGCAGATCCACATGAACAAAGAATTGCCTATGACCCCTTGATGCTGA
HDAC9V1        (1586)   AGCAGCAGATCCACATGAACAAACTGCTTTCGAAATCTATTGAACAACTG
HDAC9V2        (1586)   AGCAGCAGATCCACATGAACAAACTGCTTTCGAAATCTATTGAACAACTG
HDAC9V3        (1586)   AGCAGCAGATCCACATGAACAAACTGCTTTCGAAATCTATTGAACAACTG
CONSENSUS      (1751)   AGCAGCAGATCCACATGAACAAACTGCTTTCGAAATCTATTGAACAACTG
                                                    *SPLICE JUNCTION:
                                                     CAAA>>GAAA OR CTGC
```

```
                        1801                                              1850
BMY_HDACX_V1    (628)   AAGCAACCAGGCAGTCACCTTGAGGAAGCAGAGGAAGAGCTTCAGGGGGA
BMY_HDACX_V2   (1801)   AACACCAGTGCGTTTGTGGCAATTCCACCACCCACCCTGAGCATGCTGGA
HDAC9V1        (1636)   AAGCAACCAGGCAGTCACCTTGAGGAAGCAGAGGAAGAGCTTCAGGGGGA
HDAC9V2        (1636)   AAGCAACCAGGCAGTCACCTTGAGGAAGCAGAGGAAGAGCTTCAGGGGGA
HDAC9V3        (1636)   AAGCAACCAGGCAGTCACCTTGAGGAAGCAGAGGAAGAGCTTCAGGGGGA
CONSENSUS      (1801)   AAGCAACCAGGCAGTCACCTTGAGGAAGCAGAGGAAGAGCTTCAGGGGGA
```

```
                        1851                                              1900
BMY_HDACX_V1    (678)   CCAGGCGATGCAGGAAGACAGAGCGCCCTCTAGTGGCAACAGCACTAGGA
BMY_HDACX_V2   (1851)   CGAATACAGAGTATCTGGTCACGACTGCAAGAAACTGGGCTGCTAAATAA
HDAC9V1        (1686)   CCAGGCGATGCAGGAAGACAGAGCGCCCTCTAGTGGCAACAGCACTAGGA
HDAC9V2        (1686)   CCAGGCGATGCAGGAAGACAGAGCGCCCTCTAGTGGCAACAGCACTAGGA
HDAC9V3        (1686)   CCAGGCGATGCAGGAAGACAGAGCGCCCTCTAGTGGCAACAGCACTAGGA
CONSENSUS      (1851)   CCAGGCGATGCAGGAAGACAGAGCGCCCTCTAGTGGCAACAGCACTAGGA
```

```
                        1901                                              1950
BMY_HDACX_V1    (728)   GCGACAGCAGTGCTTGTGTGGATGACACACTGGGACAAGTTGGGGCTGTG
BMY_HDACX_V2   (1901)   ATGTGAGCGAATTCAAGGTCGAAAAGCCAGCCTCGAGGAAATACAGCTTG
HDAC9V1        (1736)   GCGACAGCAGTGCTTGTGTGGATGACACACTGGGACAAGTTGGGGCTGTG
HDAC9V2        (1736)   GCGACAGCAGTGCTTGTGTGGATGACACACTGGGACAAGTTGGGGCTGTG
HDAC9V3        (1736)   GCGACAGCAGTGCTTGTGTGGATGACACACTGGGACAAGTTGGGGCTGTG
CONSENSUS      (1901)   GCGACAGCAGTGCTTGTGTGGATGACACACTGGGACAAGTTGGGGCTGTG
```

```
                        1951                                              2000
BMY_HDACX_V1    (778)   AAGGTCAAGGAGGAACCAGTGGACAGTGATGAAGATGCTCAGATCCAGGA
BMY_HDACX_V2   (1951)   TTCATTCTGAACATCACTCACTGTTGTATGGCACCAACCCCCTGGACGGA
HDAC9V1        (1786)   AAGGTCAAGGAGGAACCAGTGGACAGTGATGAAGATGCTCAGATCCAGGA
HDAC9V2        (1786)   AAGGTCAAGGAGGAACCAGTGGACAGTGATGAAGATGCTCAGATCCAGGA
HDAC9V3        (1786)   AAGGTCAAGGAGGAACCAGTGGACAGTGATGAAGATGCTCAGATCCAGGA
CONSENSUS      (1951)   AAGGTCAAGGAGGAACCAGTGGACAGTGATGAAGATGCTCAGATCCAGGA
```

```
                        2001                                              2050
BMY_HDACX_V1    (828)   AATGGAATCTGGGGAGCAGGCTGCTTTTATGCAACAGCCTTTCCTGGAAC
BMY_HDACX_V2   (2001)   CAGAAGCTGGACCCCAGGATACTCCTAGGTGATGACTCTCAAAAGTTTTT
HDAC9V1        (1836)   AATGGAATCTGGGGAGCAGGCTGCTTTTATGCAACAGCCTTTCCTGGAAC
HDAC9V2        (1836)   AATGGAATCTGGGGAGCAGGCTGCTTTTATGCAACAGCCTTTCCTGGAAC
HDAC9V3        (1836)   AATGGAATCTGGGGAGCAGGCTGCTTTTATGCAACAGGTAATAGGCAAAG
CONSENSUS      (2001)   AATGGAATCTGGGGAGCAGGCTGCTTTTATGCAACAGCCTTTCCTGGAAC
                                                                  *SPLICE JUNCTION:
                                                                   CAG>>>CCT OR GTA
```

```
                        2051                                              2100
BMY_HDACX_V1    (878)   CCACGCACACACGTGCGCTCTCTGTGCGCCAAGCTCCGCTGGCTGCGGTT
BMY_HDACX_V2   (2051)   TTCCTCATTACCTTGTCGTGGACTTGGGTGGACAGTGACACCATTTGGA
HDAC9V1        (1886)   CCACGCACACACGTGCGCTCTCTGTGCGCCAAGCTCCGCTGGCTGCGGTT
HDAC9V2        (1886)   CCACGCACACACGTGCGCTCTCTGTGCGCCAAGCTCCGCTGGCTGCGGTT
HDAC9V3        (1886)   ATTTAGCTCCAGCATTTGTAATTAAAGTCATTATCTGAACATGAAATGCA
CONSENSUS      (2051)   CCACGCACACACGTGCGCTCTCTGTGCGCCAAGCTCCGCTGGCTGCGGTT
```

```
                        2101                                              2150
BMY_HDACX_V1    (928)   GGCATGGATGGATTAGAGAAACACCGTCTCGTCTCCAGGACTCACTCTTC
BMY_HDACX_V2   (2101)   ATGAGCTACACTCGTCCGGTGCTGCACGCATGGCTGTTGGCTGTGTCATC
HDAC9V1        (1936)   GGCATGGATGGATTAGAGAAACACCGTCTCGTCTCCAGGACTCACTCTTC
HDAC9V2        (1936)   GGCATGGATGGATTAGAGAAACACCGTCTCGTCTCCAGGACTCACTCTTC
HDAC9V3        (1936)   TTGCAGGTTTGGTAAATGGATATGATTTCCTATCAGTTTATATTTCTCTA
CONSENSUS      (2101)   GGCATGGATGGATTAGAGAAACACCGTCTCGTCTCCAGGACTCACTCTTC
```

FIG. 23E

		2151 2200
BMY_HDACX_V1	(978)	CCCTGCTGCCTCTGTTTTACCTCACCCGGCAATGGACCGCCCCCTCCAGC
BMY_HDACX_V2	(2151)	GAGCTGGCTTCCAAAGTGGCCTCAGGAGAGCTGAAGGTGAGGTCCGGGTT
HDAC9V1	(1986)	CCCTGCTGCCTCTGTTTTACCTCACCCAGCAATGGACCGCCCCCTCCAGC
HDAC9V2	(1986)	CCCTGCTGCCTCTGTTTTACCTCACCCAGCAATGGACCGCCCCCTCCAGC
HDAC9V3	(1986)	TGATTTGAGTTCAGTGTTTAAGGATTCTACCTAATGCAGATATATGTATA
CONSENSUS	(2151)	CCCTGCTGCCTCTGTTTTACCTCACCC GCAATGGACCGCCCCCTCCAGC
		2201 2250
BMY_HDACX_V1	(1028)	CTGGCTCTGCAACTGGAATTGCCTATGACCCCTTGATGCTGAAACACCAG
BMY_HDACX_V2	(2201)	GCATTAAGTGTGGGAAATCCAGAGAAGAAACTGAAACAGAGATGTTGTTA
HDAC9V1	(2036)	CTGGCTCTGCAACTGGAATTGCCTATGACCCCTTGATGCTGAAACACCAG
HDAC9V2	(2036)	CTGGCTCTGCAACTGGAATTGCCTATGACCCCTTGATGCTGAAACACCAG
HDAC9V3	(2036)	TATCTATATAGAGGTCTTTCTATATACTGATCTCTATATAGATATCAATG
CONSENSUS	(2201)	CTGGCTCTGCAACTGGAATTGCCTATGACCCCTTGATGCTGAAACACCAG
		2251 2300
BMY_HDACX_V1	(1078)	TGCGTTTGTGGCAATTCCACCACCCACCCTGAGCATGCTGGACGAATACA
BMY_HDACX_V2	(2251)	TGTGGGAATTGCGGGGAGTGTGGCGTGGTAATAAAGCGAAGGGCAGAAGG
HDAC9V1	(2086)	TGCGTTTGTGGCAATTCCACCACCCACCCTGAGCATGCTGGACGAATACA
HDAC9V2	(2086)	TGCGTTTGTGGCAATTCCACCACCCACCCTGAGCATGCTGGACGAATACA
HDAC9V3	(2086)	TTTCATTGAAAATCCACTGGTAAGGAAATACCTGTTATACTAAAATTATG
CONSENSUS	(2251)	TGCGTTTGTGGCAATTCCACCACCCACCCTGAGCATGCTGGACGAATACA
		2301 2350
BMY_HDACX_V1	(1128)	GAGTATCTGGTCACGACTGCAAGAAACTGGGCTGCTAAATAAATGTGAGC
BMY_HDACX_V2	(2301)	AAGAGGGTAGAGATGGCCACTAAGGTGTGATAATAACTCATCTGTAGGCA
HDAC9V1	(2136)	GAGTATCTGGTCACGACTGCAAGAAACTGGGCTGCTAAATAAATGTGAGC
HDAC9V2	(2136)	GAGTATCTGGTCACGACTGCAAGAAACTGGGCTGCTAAATAAATGTGAGC
HDAC9V3	(2136)	ATACATAATATTCTGAGCAGTTAATAGGCTTTAAATTTATCCCAAAGCCTG
CONSENSUS	(2301)	GAGTATCTGGTCACGACTGCAAGAAACTGGGCTGCTAAATAAATGTGAGC
		2351 2400
BMY_HDACX_V1	(1178)	GAATTCAAGGTCGAAAAGCCAGCCTGGAGGAAATACAGCTTGTTCATTCT
BMY_HDACX_V2	(2351)	GGGAGCAGCTCATCCTGCTCTCAGGGCCTTCTTCTGCCTGAGAACACTCT
HDAC9V1	(2186)	GAATTCAAGGTCGAAAAGCCAGCCTGGAGGAAATACAGCTTGTTCATTCT
HDAC9V2	(2186)	GAATTCAAGGTCGAAAAGCCAGCCTGGAGGAAATACAGCTTGTTCATTCT
HDAC9V3	(2186)	CTACACCAATTACTTCTAAAGAAAACAAATTCACTGTTATTTTGAGTTTA
CONSENSUS	(2351)	GAATTCAAGGTCGAAAAGCCAGCCTGGAGGAAATACAGCTTGTTCATTCT
		2401 2450
BMY_HDACX_V1	(1228)	GAACATCACTCACTGTTGTATGGCACCAACCCCCTGGACGGACAGAAGCT
BMY_HDACX_V2	(2401)	GCAGTCAGGGCCCACCGGTGTGCATGTAAGAGCACAGAGATAATAAGCAA
HDAC9V1	(2236)	GAACATCACTCACTGTTGTATGGCACCAACCCCCTGGACGGACAGAAGCT
HDAC9V2	(2236)	GAACATCACTCACTGTTGTATGGCACCAACCCCCTGGACGGACAGAAGCT
HDAC9V3	(2236)	TGTGTTGGAGATCAGTGACTGCTGGATAGTCTCCCAGTCTGATCAATGAAG
CONSENSUS	(2401)	GAACATCACTCACTGTTGTATGGCACCAACCCCCTGGACGGACAGAAGCT
		2451 2500
BMY_HDACX_V1	(1278)	GGACCCCAGGATACTCCTAGGTGATGACTCTCAAAAGTTTTTTTCCTCAT
BMY_HDACX_V2	(2451)	AGCTATGGTTCAGGTTAAAAATACCTTTAGTATATACATGTCTGTCATGC
HDAC9V1	(2286)	GGACCCCAGGATACTCCTAGGTGATGACTCTCAAAAGTTTTTTTCCTCAT
HDAC9V2	(2286)	GGACCCCAGGATACTCCTAGGTGATGACTCTCAAAAGTTTTTTTCCTCAT
HDAC9V3	(2286)	CATTCGATTAGTTTTTGATTTTTTGCAACATCTAGAATTTAATTTTCACA
CONSENSUS	(2451)	GGACCCCAGGATACTCCTAGGTGATGACTCTCAAAAGTTTTTTTCCTCAT
		2501 2550
BMY_HDACX_V1	(1328)	TACCTTGTGGTGGACTTGGGGTGGACAGTGACACCATTTGGAATGAGCTA
BMY_HDACX_V2	(2501)	CATGCTGAGATTCTCTTTTTGAGGCAATTTTAAAAATATGATTACTGAGAA
HDAC9V1	(2336)	TACCTTGTGGTGGACTTGGGGTGGACAGTGACACCATTTGGAATGAGCTA
HDAC9V2	(2336)	TACCTTGTGGTGGACTTGGGGTGGACAGTGACACCATTTGGAATGAGCTA
HDAC9V3	(2336)	TCACTGTACATAATGTATCATACTATAGTCTTGAACACTGTTAAAGGTAG
CONSENSUS	(2501)	TACCTTGTGGTGGACTTGGGGTGGACAGTGACACCATTTGGAATGAGCTA
		2551 2600
BMY_HDACX_V1	(1378)	CACTCGTCCGGTGCTGCACGCATGGCTGTTGGCTGTGTCATCGAGCTGGC
BMY_HDACX_V2	(2551)	GTGTGTATAAGCTCAGAATACCACCCAGAGAGAGCGAGGCAGAGAAAGGT
HDAC9V1	(2386)	CACTCGTCCGGTGCTGCACGCATGGCTGTTGGCTGTGTCATCGAGCTGGC
HDAC9V2	(2386)	CACTCGTCCGGTGCTGCACGCATGGCTGTTGGCTGTGTCATCGAGCTGGC
HDAC9V3	(2386)	TCTGGCCCGTTCCTTCCTCTCTCTTTTTTTAGTTAAGTAGAAATGTTCTGG
CONSENSUS	(2551)	CACTCGTCCGGTGCTGCACGCATGGCTGTTGGCTGTGTCATCGAGCTGGC

FIG. 23F

```
                        2601                                              2650
BMY_HDACX_V1   (1428)   TTCCAAAGTGGCCTCAGGAGAGCTGAAGAATGGGTTTGCTGTTGTGAGGC
BMY_HDACX_V2   (2601)   AAATACCAGACGGGAAGGATTGGGAGGAGGAAGGAAATTGTTGATTAGAA
     HDAC9V1   (2436)   TTCCAAAGTGGCCTCAGGAGAGCTGAAGAATGGGTTTGCTGTTGTGAGGC
     HDAC9V2   (2436)   TTCCAAAGTGGCCTCAGGAGAGCTGAAGAATGGGTTTGCTGTTGTGAGGC
     HDAC9V3   (2436)   TCACCATGCCAGTAGTCCTAGGTTATTGTGTAGGTTGCAATTGAACATAT
   CONSENSUS   (2601)   TTCCAAAGTGGCCTCAGGAGAGCTGAAGAATGGGTTTGCTGTTGTGAGGC
                        2651                                              2700
BMY_HDACX_V1   (1478)   CCCCTGGCCATCACGCTGAAGAATCCACAGCCATGGGGTTCTGCTTTTTT
BMY_HDACX_V2   (2651)   GGGTAATGATCCAGAGTGTGTTTTTCCATGAAAGAACTTAAAAAATGAGC
     HDAC9V1   (2486)   CCCCTGGCCATCACGCTGAAGAATCCACAGCCATGGGGTTCTGCTTTTTT
     HDAC9V2   (2486)   CCCCTGGCCATCACGCTGAAGAATCCACAGCCATGGGGTTCTGCTTTTTT
     HDAC9V3   (2486)   TAGGAATACAGGTGGTTTTAAATATATAGATGCAAATTGCAGCACTACTT
   CONSENSUS   (2651)   CCCCTGGCCATCACGCTGAAGAATCCACAGCCATGGGGTTCTGCTTTTTT
                        2701                                              2750
BMY_HDACX_V1   (1528)   AATTCAGTTGCAATTACCGCCAAATACTTGAGAGACCAACTAAATATAAG
BMY_HDACX_V2   (2701)   TATGCTTTATTGTTCTTTTCTTTTTATGGTCTCTTCTTTTCTACATCGTA
     HDAC9V1   (2536)   AATTCAGTTGCAATTACCGCCAAATACTTGAGAGACCAACTAAATATAAG
     HDAC9V2   (2536)   AATTCAGTTGCAATTACCGCCAAATACTTGAGAGACCAACTAAATATAAG
     HDAC9V3   (2536)   TAAATATTACATTATGTCTCACATAGCACTGCTCATTTTACTTTTATTTT
   CONSENSUS   (2701)   AATTCAGTTGCAATTACCGCCAAATACTTGAGAGACCAACTAAATATAAG
                        2751                                              2800
BMY_HDACX_V1   (1578)   CAAGATATTGATTGTAGATCTGGATGTTCACCATGGAAACGGTACCCAGC
BMY_HDACX_V2   (2751)   TGAAAAGAACAATGTCCAAACCCCAGCGTTTCCCAGTCTAAACAATTTAT
     HDAC9V1   (2586)   CAAGATATTGATTGTAGATCTGGATGTTCACCATGGAAACGGTACCCAGC
     HDAC9V2   (2586)   CAAGATATTGATTGTAGATCTGGATGTTCACCATGGAAACGGTACCCAGC
     HDAC9V3   (2586)   GTGTAATTTGATGACACTGTCTATCAAAAAAGAGCAAATGAAGCAGATGC
   CONSENSUS   (2751)   CAAGATATTGATTGTAGATCTGGATGTTCACCATGGAAACGGTACCCAGC
                        2801                                              2850
BMY_HDACX_V1   (1628)   AGGCCTTTTATGCTGACCCCAGCATCCTGTACATTTCACTCCATCGCTAT
BMY_HDACX_V2   (2801)   AAAAGCTAGAGACCTGACAGACGTTGACATTTTATTTGGTATTTTAACAG
     HDAC9V1   (2636)   AGGCCTTTTATGCTGACCCCAGCATCCTGTACATTTCACTCCATCGCTAT
     HDAC9V2   (2636)   AGGCCTTTTATGCTGACCCCAGCATCCTGTACATTTCACTCCATCGCTAT
     HDAC9V3   (2636)   AAATGTTAGTGAGAAGTAATGTGCAGCATTATGGTCCAATCAGATACAAT
   CONSENSUS   (2801)   AGGCCTTTTATGCTGACCCCAGCATCCTGTACATTTCACTCCATCGCTAT
                        2851                                              2900
BMY_HDACX_V1   (1678)   GATGAAGGGAACTTTTTCCCTGGCAGTGGAGCCCCAAATGAGGTTGGAAC
BMY_HDACX_V2   (2851)   TGCTATTTAAAGGTACGCCATGTGCGTCTTGAATGCAGTTACCCCAATAA
     HDAC9V1   (2686)   GATGAAGGGAACTTTTTCCCTGGCAGTGGAGCCCCAAATGAGGTTGGAAC
     HDAC9V2   (2686)   GATGAAGGGAACTTTTTCCCTGGCAGTGGAGCCCCAAATGAGGTTCGGTT
     HDAC9V3   (2686)   ATTGTGTCTACAATTGCAAAAAACACAGTAACAGGATGAATATTATCTGA
   CONSENSUS   (2851)   GATGAAGGGAACTTTTTCCCTGGCAGTGGAGCCCCAAATGAGGTT G A
                        2901                                              2950
BMY_HDACX_V1   (1728)   AGGCCTTGGAGAAGGGTACAATATAAATATTGCCTGGACAGGTGGCCTTG
BMY_HDACX_V2   (2901)   ACTTTGTTGGTGCTAACACGGCCTTTTAATGCACTAGTTCACACACTTCA
     HDAC9V1   (2736)   AGGCCTTGGAGAAGGGTACAATATAAATATTGCCTGGACAGGTGGCCTTG
     HDAC9V2   (2736)   TATTTCTTTAGAGCCCCACTTTTATTTGTATCTTTCAGGTAATTGCATTG
     HDAC9V3   (2736)   TATCAAGTCAAAAATCAGTTTGAAAGAAGGTGTATCATATTTTATATTGT
   CONSENSUS   (2901)   A TC   TTGAGAA     AC   TATA A ATTG CT G      T GC TTG
                        2951                                              3000
BMY_HDACX_V1   (1778)   ATCCTCCCATGGGAGATGTTGAGTACCTTGAAGCATTCAGGACCATCGTG
BMY_HDACX_V2   (2951)   TGACGCAATCTGGGTCGTGATTGATTCGGTATTTTTAGCAATTGCGGCGC
     HDAC9V1   (2786)   ATCCTCCCATGGGAGATGTTGAGTACCTTGAAGCATTCAGGACCATCGTG
     HDAC9V2   (2786)   CATGA---------------------------------------------
     HDAC9V3   (2786)   CACTAGAATCTCTTAAG-------TATAATTCCATAATGACATGGGCATA
   CONSENSUS   (2951)     CC C      GG  A         G    C    A     T     A    CGT
                        3001                                              3050
BMY_HDACX_V1   (1828)   AAGCCTGTGGCCAAAGAGTTTGATCCAGACATGGTCTTAGTATCTGCTGG
BMY_HDACX_V2   (3001)   TTAGGGAAATATATTATGACCAATAACATATGCACTGTGAGTTTTGTGAA
     HDAC9V1   (2836)   AAGCCTGTGGCCAAAGAGTTTGATCCAGACATGGTCTTAGTATCTGCTGG
     HDAC9V2   (2791)   --------------------------------------------------
     HDAC9V3   (2829)   TACCGTAACATTCTGGCAAATAACAATTAGAAAAGATAGGTTTAACAAAA
   CONSENSUS   (3001)    A C T       A  G G  T AT    A A      TT GT T TG
```

FIG. 23G

```
                     3051                                            3100
BMY_HDACX_V1  (1878) ATTTGATGCATTGGAAGGCCACACCCCTCCTCTAGGAGGGTACAAAGTGA
BMY_HDACX_V2  (3051) ACCAAGATAAAATAATTAGGATTACTTTTCTTTATGTCTAGTGAATTTTT
     HDAC9V1  (2886) ATTTGATGCATTGGAAGGCCACACCCCTCCTCTAGGAGGGTACAAAGTGA
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (2879) AAATTTACTTGTATATAATGCACCTTCAGGAGGACTATGTCCTTTGATGC
   CONSENSUS  (3051) A  T      A T  A      A  CC CT CT TA GA G    AA  TG
                     3101                                            3150
BMY_HDACX_V1  (1928) CGGCAAAATGTTTTGGTCATTTGACGAAGCAATTGATGACATTGGCTGAT
BMY_HDACX_V2  (3101) ATTCAATTACATGGCACTCTTCCAGTTGTGATTAAAAATGTGGAGTAGGA
     HDAC9V1  (2936) CGGCAAAATGTTTTGGTCATTTGACGAAGCAATTGATGACATTGGCTGAT
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (2929) TATAAAATACAAACAACT-TTGAAGGCAACAGAAGACACTGTTTATTCAA
   CONSENSUS  (3101)    CAAA    T  G     TT  A G A CA T GA       TT G TGA
                     3151                                            3200
BMY_HDACX_V1  (1978) GGACGTGTGGTGTTGGCTCTAGAAGGAGGACATGATCTCACAGCCATCTG
BMY_HDACX_V2  (3151) ATGTGCACTTCACAATGCAACGTTTGTCCAAGAAGTCTTTACTCTTAACT
     HDAC9V1  (2986) GGACGTGTGGTGTTGGCTCTAGAAGGAGGACATGATCTCACAGCCATCTG
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (2978) GTCAGTTCTTTGTCAGGTTCCTGCTGTTCTCCTACAGAAAGTGATTCTG
   CONSENSUS  (3151) G   GT      TGT  G T     G    G    AC T   TCT A     C   TCTG
                     3201                                            3250
BMY_HDACX_V1  (2028) TGATGCATCAGAAGCCTGTGTAAATGCCCTTCTAGGAAATGAGCTCGAGC
BMY_HDACX_V2  (3201) CTTTAAAGAGTCAGAGCCTACGGAAATATAATTTTGATAGCGTGAGCTCT
     HDAC9V1  (3036) TGATGCATCAGAAGCCTGTGTAAATGCCCTTCTAGGAAATGAGCTCGAGC
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3028) TGAGGGTGAACAGGAAATGCCTTGTGGAAACAGGAAGTCCAAGTGATTCA
   CONSENSUS  (3201) TGATG A  A AAG     T    ATG        T  GA A GAG  G
                     3251                                            3300
BMY_HDACX_V1  (2078) CACTTGCAGAAGATATTCTCCACCAAAGCCCGAATATGAATGCTGTTATT
BMY_HDACX_V2  (3251) ATTTAAAAAGTAGATGTGCCTGTATATATTTGACATAAGTAGTATTAGGA
     HDAC9V1  (3086) CACTTGCAGAAGATATTCTCCACCAAAGCCCGAATATGAATGCTGTTATT
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3078) TGTACTGAGGAATGTAGGAAAAAAAATCTGAGGATAGTGCTTTACTCTTT
   CONSENSUS  (3251)    T  AG A     T  C A  AA     GAATA    TG    T  TT
                     3301                                            3350
BMY_HDACX_V1  (2128) TCTTTACAGAAGATCATTGAAATTCAAAGCAAGTATTGGAAGTCAGTAAG
BMY_HDACX_V2  (3301) CATTGCTCATCTCAGGGGATATATGGGGTCATTAATGTGGTGCTTACTCT
     HDAC9V1  (3136) TCTTTACAGAAGATCATTGAAATTCAAAGTATGTCTTTAAAGTTCTCTTA
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3128) CTGTTTTTAAAGGGCACTCTATGAATTGATTTATTGTCTAAGAAAATAAC
   CONSENSUS  (3301)   TTT     AAG  CA T  A  T       AT T TT  AAG
                     3351                                            3400
BMY_HDACX_V1  (2178) GATGGTGGCTGTGCCAAGGGGCTGTGCTCTGGCTGGTGCTCAGTTGCAAG
BMY_HDACX_V2  (3351) TCAGTCTTTACCTTTGAAAATGAGCAAAAAAAAAAAAAAAA---------
     HDAC9V1  (3186) A-------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3178) ACCACAAGTAGGGAAATTGTTACGGAAGCTTTTCACTGGAACATTTCCTT
   CONSENSUS  (3351)                        G
                     3401                                            3450
BMY_HDACX_V1  (2228) AGGAGACAGAGACCGTTTCTGCCCTGGCCTCCCTAACAGTGGATGTGGAA
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3228) CATATTCCCTTTTGATATGTTTACCTTGTTTTATAGGTTTACTTTTGTTA
   CONSENSUS  (3401)
                     3451                                            3500
BMY_HDACX_V1  (2278) CAGCCCTTTGCTCAGGAAGACAGCAGAACTGCTGGTGAGCCTATGGAAGA
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3278) AGCTAGTTAAAGGTTCGTTGTATTAAGACCCCTTTAATATGGATAATCCA
   CONSENSUS  (3451)
```

FIG. 23H

```
                        3501                                              3550
BMY_HDACX_V1  (2328) GGAGCCAGCCTTGTGAAGTGCCAAGTCCCCCTCTGATATTTCCTGTGTGT
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3328) AATTGACCTAGAATCTTTGTGAGGTTTTTTCTATTAAAATATTTATATTT
   CONSENSUS  (3501)
                        3551                                              3600
BMY_HDACX_V1  (2378) GACATCATTGTGTATCCCCCCACCCCAGTACCCTCAGACATGTCTTGTCT
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3378) CTAAATCCGAGGTATTTCAAGGTGTAGTATCCTATTTCAAAGGAGATATA
   CONSENSUS  (3551)
                        3601                                              3650
BMY_HDACX_V1  (2428) GCTGCCTGGGTGGCACAGATTCAATGGAACATAAACACTGGGCACAAAAT
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3428) GCAGTTTTGCCAAATGTAGACATTGTTCAACTGTATGTTATTGGCACGTG
   CONSENSUS  (3601)
                        3651                                              3700
BMY_HDACX_V1  (2478) TCTGAACAGCAGCTTCACTTGTTCTTTGGATGGACTTGAAAGGGCATTAA
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3478) TTGTTTACATTTTGCTGTGACATTTAAAAATATTTCTTTAAAAATGTTAC
   CONSENSUS  (3651)
                        3701                                              3750
BMY_HDACX_V1  (2528) AGATTCCTTAAACGTAACCGCTGTGATTCTAGAGTTACAGTAAACCACGA
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3528) TGCTAAAGATACATTATCCTTTTTTAAAAAGTCTCCATTCAAATTAAATT
   CONSENSUS  (3701)
                        3751                                              3800
BMY_HDACX_V1  (2578) TTGGAAGAAACTGCTTCCAGCATGCTTTTAATATGCTGGGTGACCCACTC
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3578) AACATAACTAGAAGTTAGAAAGTTTAAAAGTTTTCCACATAATGAAAGTC
   CONSENSUS  (3751)
                        3801                                              3850
BMY_HDACX_V1  (2628) CTAGACACCAAGTTTGAACTAGAAACATTCAGTACAGCACTAGATATTGT
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3628) CTTCTGATAATTTGACAAATAGCTATAATAGGAACACTCCCTATCACCAA
   CONSENSUS  (3801)
                        3851                                              3900
BMY_HDACX_V1  (2678) TAATTTCAGAAGCTATGACAGCCAGTGAAATTTTGGGCAAAACCTGAGAC
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3678) CATATTTTGGTTAGTATATTCCTTCATATTAAAATGACTTTTTGTCAGTT
   CONSENSUS  (3851)
                        3901                                              3950
BMY_HDACX_V1  (2728) ATAGTCATTCCTGACATTCTGATCAGCTTTTTTTGGGGTAATTTGTTTTT
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3728) GTTTTGCATTAAAAATATGGCATGCCTAAGATAAAATTGTATATTTTTTC
   CONSENSUS  (3901)
```

FIG. 23I

```
                      3951                                              4000
BMY_HDACX_V1  (2778) CAAACAGTCTTAACTTGTTTACAAGATTTGCTTTTAGCTATGAACGGATC
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3778) CATCTCATAAATATTCATTTTCTTCAAAGTCTTTTTTCAATCTCATAAAA
   CONSENSUS  (3951)
                      4001                                              4050
BMY_HDACX_V1  (2828) GTAATTCCACCCAGAATGTAATGTTTCTTGTTTGTTTGTTTTGTTTTGTT
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3828) AAGGGATAGTGCATCTTTTAAAATACATTTTATTTGGGGAGGAACATGTG
   CONSENSUS  (4001)
                      4051                                              4100
BMY_HDACX_V1  (2878) AGGGTTTTTTTCTCAACTTTAACACACAGTTCAACTGTTCCTAGTAAAAG
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3878) GCTGAGCAGACTTTTGTATAATATTACTTCAAAGATATGTAATCACAAAC
   CONSENSUS  (4051)
                      4101                                              4150
BMY_HDACX_V1  (2928) TTCAAGATGGAGGAACTAGCATGAGGCTTTTTTCAGTATCTCGAAGTCCA
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3928) AAAAAAAACTATTTTTTATAATGTCATTTGAGAGAGTTTCATCAGTACAG
   CONSENSUS  (4101)
                      4151                                              4200
BMY_HDACX_V1  (2978) AATGCCAAAGGAACCTCACACACTGTTTGTAATGGTGCAATATTTTATAT
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (3978) TTGGTGGACGTTAATTGTTTGAATTTGATAGTCTTTGAATTTAATCAAGA
   CONSENSUS  (4151)
                      4201                                              4250
BMY_HDACX_V1  (3028) CACTTTTTTTTAAACATCCCCAACATCTTTGTGTTCTCACACACAGGCAA
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (4028) AACTACCTGGAACCAGTGAAAAGGAAAGCTGGACTTAAATAATCTTAGAA
   CONSENSUS  (4201)
                      4251                                              4300
BMY_HDACX_V1  (3078) TTTGCAATGTTGCAATTGTGTTGGAGAATGAAGTCCCCCCACCTCCCAGC
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (4078) TTAATTGATAAATGTCTCTTTTAAAATCTACTGTATTTATTATAATTTAC
   CONSENSUS  (4251)
                      4301                                              4350
BMY_HDACX_V1  (3128) CACACACACATCCTTTGTTCTCATGACAGTAGGTCTGAGCAAATGTTCCA
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (4128) ACCCTTGAAGGTGATCTCTTGTTTTGTGTTGTAAATATATTGTTTGTATG
   CONSENSUS  (4301)
                      4351                                              4400
BMY_HDACX_V1  (3178) CCAAGCATTTTCAGTGTCTTTGAAAAGCACGTAACTTTTCAAAGGTGGTC
BMY_HDACX_V2  (3392) --------------------------------------------------
     HDAC9V1  (3187) --------------------------------------------------
     HDAC9V2  (2791) --------------------------------------------------
     HDAC9V3  (4178) TTTCCCTTCTTGCCTTCTGTTATAAGTCTCTTCCTTTCTCAAATAAAGTT
   CONSENSUS  (4351)
```

FIG. 23J

```
                      4401                                              4450
BMY_HDACX_V1  (3228)  TTAATTTGCTGCATATCTATCAAGGACTTATTCACTCACCTTTCCTTTTC
BMY_HDACX_V2  (3392)  --------------------------------------------------
     HDAC9V1  (3187)  --------------------------------------------------
     HDAC9V2  (2791)  --------------------------------------------------
     HDAC9V3  (4228)  TTTTTTAAAAG---------------------------------------
   CONSENSUS  (4401)
                      4451                                              4500
BMY_HDACX_V1  (3278)  TGCCCTCTATCAATTGATTTCTTCTTACCTTTCATCATTCATTCCTTCCT
BMY_HDACX_V2  (3392)  --------------------------------------------------
     HDAC9V1  (3187)  --------------------------------------------------
     HDAC9V2  (2791)  --------------------------------------------------
     HDAC9V3  (4239)  --------------------------------------------------
   CONSENSUS  (4451)
                      4501                                              4550
BMY_HDACX_V1  (3328)  TTAGAAAAACTGAAGATTACCCATAATCTCCTCTTATTACTTGAGGGCCT
BMY_HDACX_V2  (3392)  --------------------------------------------------
     HDAC9V1  (3187)  --------------------------------------------------
     HDAC9V2  (2791)  --------------------------------------------------
     HDAC9V3  (4239)  --------------------------------------------------
   CONSENSUS  (4501)
                      4551                                              4600
BMY_HDACX_V1  (3378)  TGACTATTTAGTTTATTTTGTTTACTTTACAGGTTAACACAGTTGTTTTG
BMY_HDACX_V2  (3392)  --------------------------------------------------
     HDAC9V1  (3187)  --------------------------------------------------
     HDAC9V2  (2791)  --------------------------------------------------
     HDAC9V3  (4239)  --------------------------------------------------
   CONSENSUS  (4551)
                      4601                                              4650
BMY_HDACX_V1  (3428)  TCTGATTGCATTTTATTAACTGTGAAGCCGTTGAAATGAATATCACTTAA
BMY_HDACX_V2  (3392)  --------------------------------------------------
     HDAC9V1  (3187)  --------------------------------------------------
     HDAC9V2  (2791)  --------------------------------------------------
     HDAC9V3  (4239)  --------------------------------------------------
   CONSENSUS  (4601)
                      4651                                              4700
BMY_HDACX_V1  (3478)  GCAACGTTGCTAAATTTCTATGTGTTTGAAATGTGTTAATGAAGGCACTG
BMY_HDACX_V2  (3392)  --------------------------------------------------
     HDAC9V1  (3187)  --------------------------------------------------
     HDAC9V2  (2791)  --------------------------------------------------
     HDAC9V3  (4239)  --------------------------------------------------
   CONSENSUS  (4651)
                      4701                                              4750
BMY_HDACX_V1  (3528)  CTTATTTGTAGTCACCTTGAACTGACTTAACCTAGAAGCTGTGCCTTCTT
BMY_HDACX_V2  (3392)  --------------------------------------------------
     HDAC9V1  (3187)  --------------------------------------------------
     HDAC9V2  (2791)  --------------------------------------------------
     HDAC9V3  (4239)  --------------------------------------------------
   CONSENSUS  (4701)
                      4751                                              4800
BMY_HDACX_V1  (3578)  GTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
BMY_HDACX_V2  (3392)  --------------------------------------------------
     HDAC9V1  (3187)  --------------------------------------------------
     HDAC9V2  (2791)  --------------------------------------------------
     HDAC9V3  (4239)  --------------------------------------------------
   CONSENSUS  (4751)
                      4801                  4823
BMY_HDACX_V1  (3628)  AAAAAAAAAAAAAAAAAAAAAAA
BMY_HDACX_V2  (3392)  -----------------------
     HDAC9V1  (3187)  -----------------------
     HDAC9V2  (2791)  -----------------------
     HDAC9V3  (4239)  -----------------------
   CONSENSUS  (4801)
```

FIG. 23K

```
                       1                                                50
     HDAC9V2    (1)  ------------------------MHSMISSVDVKSEVPVGLEPIS---P
     HDAC9V1    (1)  ------------------------MHSMISSVDVKSEVPVGLEPIS---P
     HDAC9V3    (1)  ------------------------MHSMISSVDVKSEVPVGLEPIS---P
 BMY_HDACX_V1   (1)  --------------------------------------------------
 BMY_HDACX_V2   (1)  ------------------------MHSMISSVDVKSEVPVGLEPIS---P
        HDA5    (1)  MNSPNESDGMSGREPSLEILPRTSLHSIPVTVEVKPVLPRAMPSSMGGGG
        HDA4    (1)  MSSQSHPDGLSGRDQPVELLNPARVNHMPSTVDVATALPLQVAPSA--VP
   CONSENSUS    (1)  M S     DGLSGRD  LEIL     M  M  SVDV    VP  L     GG
                     51                                                100
     HDAC9V2   (24)  LDLRTDLRMMMP---VVDPVVREKQLQQELLLIQQQQQIQKQLLIAEFQK
     HDAC9V1   (24)  LDLRTDLRMMMP---VVDPVVREKQLQQELLLIQQQQQIQKQLLIAEFQK
     HDAC9V3   (24)  LDLRTDLRMMMP---VVDPVVREKQLQQELLLIQQQQQIQKQLLIAEFQK
 BMY_HDACX_V1   (1)  --------------------------------------------------
 BMY_HDACX_V2  (24)  LDLRTDLRMMMP---VVDPVVREKQLQQELLLIQQQQQIQKQLLIAEFQK
        HDA5   (51)  GGSPSPVELRGALVGSVDPTLREQQLQQELLALKQQQQLQKQLLFAEFQK
        HDA4   (49)  MDLRLDHQFSLP---VAEPALREQQLQQELLALKQKQQIQRQILIAEFQR
   CONSENSUS   (51)                 LVG  DP VRE QLQQELL I Q QQIQKQLL AEFQK
                     101                                               150
     HDAC9V2   (71)  QHENLTRQHQAQLQEHIK---ELLAIKQQQELLEKEQK--LEQQRQEQ--
     HDAC9V1   (71)  QHENLTRQHQAQLQEHIK---ELLAIKQQQELLEKEQK--LEQQRQEQ--
     HDAC9V3   (71)  QHENLTRQHQAQLQEHIK---ELLAIKQQQELLEKEQK--LEQQRQEQ--
 BMY_HDACX_V1   (1)  --------------------------------------------------
 BMY_HDACX_V2  (71)  QHENLTRQHQAQLQEHIKLQQELLAIKQQQELLEKEQK--LEQQRQEQ--
        HDA5  (101)  QHDHLTRQHEVQLQKHLKQQQEMLAAKQQQEMLAAKRQQELEQQRQREQQ
        HDA4   (96)  QHEQLSRQHEAQLHEHIKQQQEMLAMKHQQELLEHQRK--LERHRQEQ--
   CONSENSUS  (101)  QHE LTRQH  QL  HIK QQELLA K QQELL       QELE  RQ  QQ
                     151                                               200
     HDAC9V2  (114)  ---EVERHRREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDT
     HDAC9V1  (114)  ---EVERHRREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDT
     HDAC9V3  (114)  ---EVERHRREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDT
 BMY_HDACX_V1   (1)  --------------------------------------------------
 BMY_HDACX_V2 (117)  ---EVERHRREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDT
        HDA5  (151)  RQEELEKQRLEQQLLILRNKEKSKESAIASTEVKLRLQEFLLSKSKEPTP
        HDA4  (142)  ---ELEKQHREQKLQQLKNKEKGKESAVASTEVKMKLQEFVLNKKKALAH
   CONSENSUS  (151)  RQEEVER    EQ L    LR KDR RE AVASTEVK KLQEFLL K
                     201                                               250
     HDAC9V2  (161)  PTNGKNHSVSRHPKLWYTAAHHTSLDQSSPPLS---GTSPSYKYTLPGAQ
     HDAC9V1  (161)  PTNGKNHSVSRHPKLWYTAAHHTSLDQSSPPLS---GTSPSYKYTLPGAQ
     HDAC9V3  (161)  PTNGKNHSVSRHPKLWYTAAHHTSLDQSSPPLS---GTSPSYKYTLPGAQ
 BMY_HDACX_V1   (1)  --------------------------------------------------
 BMY_HDACX_V2 (164)  PTNGKNHSVSRHPKLWYTAAHHTSLDQSSPPLS---GTSPSYKYTLPGAQ
        HDA5  (201)  GG--LNHSLPQHPKCW-G-AHHASLDQSSPPQSGPPGTPPSYKLPLPGPY
        HDA4  (189)  RN--LNHCISSDPRYWYGKTQHSSLDQSSPPQS---GVSTSYNHPVLGMY
   CONSENSUS  (201)    NG NH V     PK WY     H SLDQSSPP SGPPG    SY    L G
                     251                                               300
     HDAC9V2  (208)  DAKDDFPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGNVVTSFKKR
     HDAC9V1  (208)  DAKDDFPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGNVVTSFKKR
     HDAC9V3  (208)  DAKDDFPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGNVVTSFKKR
 BMY_HDACX_V1   (1)  --------------------------------------------------
 BMY_HDACX_V2 (211)  DAKDDFPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGNVVTSFKKR
        HDA5  (247)  DSRDDFPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGTVISTFKKR
        HDA4  (234)  DAKDDFPLRKTASEPNLKLRSRLKQKVAERRSSPLLRRKDGPVVTALKKR
   CONSENSUS  (251)  DAKDDFPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDG VVT   KKR
                     301                                               350
     HDAC9V2  (258)  MFEVT-----ESSVSSSSPGSGPSSPNNGPTGSVTENETSVLPPTPHAEQ
     HDAC9V1  (258)  MFEVT-----ESSVSSSSPGSGPSSPNNGPTGSVTENETSVLPPTPHAEQ
     HDAC9V3  (258)  MFEVT-----ESSVSSSSPGSGPSSPNNGPTGSVTENETSVLPPTPHAEQ
 BMY_HDACX_V1   (1)  ----------------------------------AENETSVLPPTPHAEQ
 BMY_HDACX_V2 (261)  MFEVT-----ESSVSSSSPGSGPSSPNNGPTGSVTENETSVLPPTPHAEQ
        HDA5  (297)  AVEITGAGPGASSVCNSAPGSGPSSPN-SSHSTIAENGFTGSVPNIPTEM
        HDA4  (284)  PLDVT------DSACSSAPGSGPSSPNNSSGSVSAENGIAPAVPSIPAET
   CONSENSUS  (301)    EVTGAGPG  S    SSPGSGPSSPNN          EN        P     E
```

FIG. 24A

```
                    351                                                400
HDAC9V2       (303) MVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQLNASNSLK----
HDAC9V1       (303) MVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQLNASNSLK----
HDAC9V3       (303) MVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQLNASNSLK----
BMY_HDACX_V1   (17) MVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQLNASNSLK----
BMY_HDACX_V2  (306) MVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQLNASNSLK----
HDA5          (346) LPQHRALPLDSSPNQFSLYTSPSLPNISLGLQATVTVTNSHLTASPKLST
HDA4          (328) SLAHRLVAREGSAAPLPLYTSPSLPNITLGLPATGPSAGTAG--------
CONSENSUS     (351)           I         T                       S  KLST
                    401                                                450
HDAC9V2       (349) --EKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALL
HDAC9V1       (349) --EKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALL
HDAC9V3       (349) --EKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALL
BMY_HDACX_V1   (63) --EKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALL
BMY_HDACX_V2  (352) --EKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALL
HDA5          (396) QQEAERQALQSLRQGGTLTSKFMSTSSIPGCLLGVALEGDGSPHGHASLL
HDA4          (370) QQDTERLTLPALQQR-----LSLFPGTHLTPYLSTSPLERDGGAAHSPLL
CONSENSUS     (401) QQE  K      L Q   L G                         H  LL
                    451                                                500
HDAC9V2       (397) QHLLLKEQMRQQKLLV--AGGVPLHPQSPLATKERISPGIRGTHKLPRHR
HDAC9V1       (397) QHLLLKEQMRQQKLLV--AGGVPLHPQSPLATKERISPGIRGTHKLPRHR
HDAC9V3       (397) QHLLLKEQMRQQKLLV--AGGVPLHPQSPLATKERISPGIRGTHKLPRHR
BMY_HDACX_V1  (111) QHLLLKEQMRQQKLLV--AGGVPLHPQSPLATKERISPGIRGTHKLPRHR
BMY_HDACX_V2  (400) QHLLLKEQMRQQKLLV--AGGVPLHPQSPLATKERISPGIRGTHKLPRHR
HDA5          (446) QHVLLLEQARQQSTLI----AVPLHGQSPLVTGERVATSMRTVGKLPRHR
HDA4          (415) QHMVLLEQPPAQAPLVTGLQALPLHAQS-LVGADRVSPSI---HKLRQHR
CONSENSUS     (451) QHLLL EQ    Q   LVTG GGVPLH QSPL   ERIS   IR   KL  HR
                    501                                                550
HDAC9V2       (445) PLNRTQSAPLPQS--TLAQLVIQQQHQQFLEKQKQ-Y-QQQIHMNKLLSK
HDAC9V1       (445) PLNRTQSAPLPQS--TLAQLVIQQQHQQFLEKQKQ-Y-QQQIHMNKLLSK
HDAC9V3       (445) PLNRTQSAPLPQS--TLAQLVIQQQHQQFLEKQKQ-Y-QQQIHMNKLLSK
BMY_HDACX_V1  (159) PLNRTQSAPLPQS--TLAQLVIQQQHQQFLEKQKQ-Y-QQQIHMNKLLSK
BMY_HDACX_V2  (448) PLNRTQSAPLPQS--TLAQLVIQQQHQQFLEKQKQ-Y-QQQIHMNKELPM
HDA5          (492) PLSRTQSSPLPQSPQALQQLVMQQQHQQFLEKQKQ----QQLQLGKILTK
HDA4          (461) PLGRTQSAPLPQNAQALQHLVIQQQHQQFLEKHKQQFQQQQLQMNKIIPK
CONSENSUS     (501) PL RTQSAPLPQ   Q L  LVIQQQHQQFLEK KQQYQQQQI M K L
                    551                                                600
HDAC9V2       (491) SIEQLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDSSACVDDTLG
HDAC9V1       (491) SIEQLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDSSACVDDTLG
HDAC9V3       (491) SIEQLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDSSACVDDTLG
BMY_HDACX_V1  (205) SIEQLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDSSACVDDTLG
BMY_HDACX_V2  (494) TP------------------------------------------------
HDA5          (538) TGELPRQPTTHPEETEEELTEQQEVLLGEGALTMPREGSTESESTQEDLE
HDA4          (511) PSEPARQPESHPEETEEELREHQ-ALLDEPYLDRLPGQKEAHAQAGVQVK
CONSENSUS     (551)    E  KQP SH EE EEEL    Q                          L
                    601                                                650
HDAC9V2       (541) QVGAVKVKEEP-------VDSDEDAQIQEMESGEQAAFMQQPFLEPTHTR
HDAC9V1       (541) QVGAVKVKEEP-------VDSDEDAQIQEMESGEQAAFMQQPFLEPTHTR
HDAC9V3       (541) QVGAVKVKEEP-------VDSDEDAQIQEMESGEQAAFMQQVIGKDLAPG
BMY_HDACX_V1  (255) QVGAVKVKEEP-------VDSDEDAQIQEMESGEQAAFMQQPFLEPTHTR
BMY_HDACX_V2  (496) --------------------------------------------------
HDA5          (588) EEDEEEDGEEEEDCIQVKDEEGESGAEEGPDLEEPGAGYKKLF-SDAQPL
HDA4          (560) QEPIESDEEEAE-------PPREVEPGQRQPSEQELLFRQQALLLEQQRI
CONSENSUS     (601)           EE EDCIQVK     E
                    651                                                700
HDAC9V2       (584) ALSVR-QAPLAAVGMD-GLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGS
HDAC9V1       (584) ALSVR-QAPLAAVGMD-GLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGS
HDAC9V3       (584) FVIKVII-------------------------------------------
BMY_HDACX_V1  (298) ALSVR-QAPLAAVGMD-GLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGS
BMY_HDACX_V2  (496) --------------------------------------------------
HDA5          (637) QPLQVYQAPLSLATVP-----HQALGRTQSSPAAPGGMKSPPDQPVKHLF
HDA4          (603) HQLRNYQASMEAAGIPVSFGGHRPLSRAQSSPASATFPVSVQEPPTKPRF
CONSENSUS     (651)        A L    M V     H   V R  SSPAA          D P
```

FIG. 24B

```
                        701                                               750
HDAC9V2      (632) ATGIAYDPLMLKHQCVCGNSTTHPEHAGRIQSIWSRLQETGLLNKCERIQ
HDAC9V1      (632) ATGIAYDPLMLKHQCVCGNSTTHPEHAGRIQSIWSRLQETGLLNKCERIQ
HDAC9V3      (591) --------------------------------------------------
BMY_HDACX_V1 (346) ATGIAYDPLMLKHQCVCGNSTTHPEHAGRIQSIWSRLQETGLLNKCERIQ
BMY_HDACX_V2 (496) --------------------------------------------------
HDA5         (682) TTGVVYDTFMLKHQCMCGNTHVHPEHAGRIQSIWSRLQETGLLSKCERIR
HDA4         (653) TTGLVYDTLMLKHQCTCGSSSSHPEHAGRIQSIWSRLQETGLRGKCECIR
CONSENSUS    (701)  TGI YD  MLKHQC CG S  HPEHAGRIQSIWSRLQETGL  KCE I
                   <--        HISTONE DEACETYLASE MOTIF (PF00850)      ->
                   751                                               800
HDAC9V2      (682) GRKASLEEIQLVHSEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLPC
HDAC9V1      (682) GRKASLEEIQLVHSEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLPC
HDAC9V3      (591) --------------------------------------------------
BMY_HDACX_V1 (396) GRKASLEEIQLVHSEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLPC
BMY_HDACX_V2 (496) --------------------------------------------------
HDA5         (732) GRKATLDEIQTVHSEYHTLLYGTSPLNRQKLDSKKLLGPISQKMYAVLPC
HDA4         (703) GRKATLEELQTVHSEAHTLLYGTNPLNRQKLDSKKLLG-SLASVFVRLPC
CONSENSUS    (751) GRKASLEEIQ VHSE HSLLYGT PL  QKLD R LLG       F  LPC
                   <--        HISTONE DEACETYLASE MOTIF (PF00850)      ->
                   801                                               850
HDAC9V2      (732) GGLGVDSDTIWNELHSSGAARMAVGCVIELASKVASGELKNGFAVVRPPG
HDAC9V1      (732) GGLGVDSDTIWNELHSSGAARMAVGCVIELASKVASGELKNGFAVVRPPG
HDAC9V3      (591) --------------------------------------------------
BMY_HDACX_V1 (446) GGLGVDSDTIWNELHSSGAARMAVGCVIELASKVASGELKNGFAVVRPPG
BMY_HDACX_V2 (496) --------------------------------------------------
HDA5         (782) GGIGVDSDTVWNEMHSSSAVRMAVGCLLELAFKVAAGELKNGFAIIRPPG
HDA4         (752) GGVGVDSDTIWNEVHSAGAARLAVGCVVELVFKVATGELKNGFAVVRPPG
CONSENSUS    (801) GGLGVDSDTIWNELHSS A RMAVGCVIEL  KVA GELKNGFAVVRPPG
                    <--        HISTONE DEACETYLASE MOTIF (PF00850)      ->
                   851                                               900
HDAC9V2      (782) HHAEESTAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNGTQQAF
HDAC9V1      (782) HHAEESTAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNGTQQAF
HDAC9V3      (591) --------------------------------------------------
BMY_HDACX_V1 (496) HHAEESTAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNGTQQAF
BMY_HDACX_V2 (496) --------------------------------------------------
HDA5         (832) HHAEESTAMGFCFFNSVAITAKLLQQKLNVGKVLIVDWDIHHGNGTQQAF
HDA4         (802) HHAEESTPMGFCYFNSVAVAAKLLQQRLSVSKILIVDWDVHHGNGTQQAF
CONSENSUS    (851) HHAEEST MGFCFFNSVAI AK L    L I KILIVD DVHHGNGTQQAF
                    <--        HISTONE DEACETYLASE MOTIF (PF00850)      ->
                   901                                               950
HDAC9V2      (832) YADPSILYISLHRYDEGNFFPGSGAPNEVRFISLEPHFYLYLSGNCIA--
HDAC9V1      (832) YADPSILYISLHRYDEGNFFPGSGAPNEVGTGLGEGYNINIAWTGGLDPP
HDAC9V3      (591) --------------------------------------------------
BMY_HDACX_V1 (546) YADPSILYISLHRYDEGNFFPGSGAPNEVGTGLGEGYNINIAWTGGLDPP
BMY_HDACX_V2 (496) --------------------------------------------------
HDA5         (882) YNDPSVLYISLHRYDNGNFFPGSGAPEEVGGGPGVGYNVNVAWTGGVDPP
HDA4         (852) YSDPSVLYMSLHRYDDGNFFPGSGAPDEVGTGPGVGFNVNMAFTGGLDPP
CONSENSUS    (901) Y DPSILYISLHRYD GNFFPGSGAP EV                   L PP
                   <--        HISTONE DEACETYLASE MOTIF (PF00850)      ->
                   951                                              1000
HDAC9V2      (880) --------------------------------------------------
HDAC9V1      (882) MGDVEYLEAFRTIVKPVAKEFDPDMVLVSAGFDALEGHTPPLGGYKVTAK
HDAC9V3      (591) --------------------------------------------------
BMY_HDACX_V1 (596) MGDVEYLEAFRTIVKPVAKEFDPDMVLVSAGFDALEGHTPPLGGYKVTAK
BMY_HDACX_V2 (496) --------------------------------------------------
HDA5         (932) IGDVEYLTAFRTVVMPIAHEFSPDVVLVSAGFDAVEGHLSPLGGYSVTAR
HDA4         (902) MGDAEYLAAFRTVVMPIASEFAPDVVLVSSGFDAVEGHPTPLGGYNLSAR
CONSENSUS    (951) MGD EYL AFRTIV PIA EF PDMVLVSAGFDALEGH  PLGGY VTAK
                   <-- .      HISTONE DEACETYLASE MOTIF (PF00850)      ->
```

FIG. 24C

```
                        1001                                              1050
HDAC9V2         (880)   --------------------------------------------------
HDAC9V1         (932)   CFGHLTKQLMTLADGRVVLALEGGHDLTAICDASEACVNALLGNELEPLA
HDAC9V3         (591)   --------------------------------------------------
BMY_HDACX_V1    (646)   CFGHLTKQLMTLADGRVVLALEGGHDLTAICDASEACVNALLGNELEPLA
BMY_HDACX_V2    (496)   --------------------------------------------------
HDA5            (982)   CFGHLTRQLMTLAGGRVVLALEGGHDLTAICDASEACVSALLSVELQPLD
HDA4            (952)   CFGYLTKQLMGLAGGRIVLALEGGHDLTAICDASEACVSALLGNELDPLP
CONSENSUS      (1001)   CFGHLTKQLM LA GRVVLALEGGHDLTAICDASEACV ALL  EL PL
                        <--       HISTONE DEACETYLASE MOTIF (PF00850) ->

                        1051                                              1100
HDAC9V2         (880)   --------------------------------------------------
HDAC9V1         (982)   EDILHQSPNMNAVISLQKIIEIQSMSLKFS--------------------
HDAC9V3         (591)   --------------------------------------------------
BMY_HDACX_V1    (696)   EDILHQSPNMNAVISLQKIIEIQSKYWKSVRMVAVPRGCAL--AGAQLQE
BMY_HDACX_V2    (496)   --------------------------------------------------
HDA5           (1032)   EAVLQQKPNINAVATLEKVIEIQSKHWSCVQKFAAGLGRSLREAQAGETE
HDA4           (1002)   EKVLQQRPNANAVRSMEKVMEIHSKYWRCLQRTTSTAGRSLIEAQTCENE
CONSENSUS      (1051)   E IL Q PN NAV SL KIIEI S                 G SL EA     E
                        1101                                    1141
HDAC9V2         (880)   -----------------------------------------
HDAC9V1        (1012)   -----------------------------------------
HDAC9V3         (591)   -----------------------------------------
BMY_HDACX_V1    (744)   ETETVSAL----ASLTVDVEQPFAQEDSRTAGEPMEEEPAL
BMY_HDACX_V2    (496)   -----------------------------------------
HDA5           (1082)   EAETVSAMALLSVGAEQAQAAAAREHSPRPAEEPMEQEPAL
HDA4           (1052)   EAETVTAMASLSVGVKPAEK----RP----DEEPMEEEPPL
CONSENSUS      (1101)   E ETVSAMA LS                  R   EPME EP L
```

FIG. 24D

```
BMY_HDAL1   ------------------------------------------------------------
BMY_HDAL2   ------------------------------------------------------------
BMY_HDAL3   ------------------------------------------------------------
HDAC9C      MHSMISSVDVKSEVPVGLEPISPLDLRTDLRMMMPVVDPVVREKQLQQELLLIQQQQQIQ
HDACX_V1    ------------------------------------------------------------
HDACX_V2    MHSMISSVDVKSEVPVGLEPISPLDLRTDLRMMMPVVDPVVREKQLQQELLLIQQQQQIQ

BMY_HDAL1   ------------------------------------------------------------
BMY_HDAL2   ------------------------------------------------------------
BMY_HDAL3   ------------------------------------------------------------
HDAC9C      KQLLIAEFQKQHENLTRQHQAQLQEHIKLQQELLAIKQQQELLEKEQKLEQQRQEQEVER
HDACX_V1    ------------------------------------------------------------
HDACX_V2    KQLLIAEFQKQHENLTRQHQAQLQEHIKLQQELLAIKQQQELLEKEQKLEQQRQEQEVER

BMY_HDAL1   ------------------------------------------------------------
BMY_HDAL2   ------------------------------------------------------------
BMY_HDAL3   ------------------------------------------------------------
HDAC9C      HRREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDTPTNGKNHSVSRHPKLWY
HDACX_V1    ------------------------------------------------------------
HDACX_V2    HRREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDTPTNGKNHSVSRHPKLWY

BMY_HDAL1   ------------------------------------------------------------
BMY_HDAL2   ------------------------------------------------------------
BMY_HDAL3   ------------------------------------------------------------
HDAC9C      TAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKDDFPLRKTASEPNLKVRSRLKQKVAER
HDACX_V1    ------------------------------------------------------------
HDACX_V2    TAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKDDFPLRKTASEPNLKVRSRLKQKVAER

BMY_HDAL1   ------------------------------------------------------------
BMY_HDAL2   ------------------------------------------------------------
BMY_HDAL3   ------------------------------------------------------------
HDAC9C      RSSPLLRRKDGNVVTSFKKRMFEVTESSVSSSSPGSGPSSPNNGPTGSVTENETSVLPPT
HDACX_V1    -------------------------------------------------AENETSVLPPT
HDACX_V2    RSSPLLRRKDGNVVTSFKKRMFEVTESSVSSSSPGSGPSSPNNGPTGSVTENETSVLPPT

BMY_HDAL1   ------------------------------------------------------------
BMY_HDAL2   ------------------------------------------------------------
BMY_HDAL3   ------------------------------------------------------------
HDAC9C      PHAEQMVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQLNASNSLKEKQKCETQT
HDACX_V1    PHAEQMVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQLNASNSLKEKQKCETQT
HDACX_V2    PHAEQMVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQLNASNSLKEKQKCETQT

BMY_HDAL1   ------------------------------------------------------------
BMY_HDAL2   ------------------------------------------------------------
BMY_HDAL3   ------------------------------------------------------------
HDAC9C      LRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALLQHLLLKEQMRQQKLLVAGGVP
HDACX_V1    LRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALLQHLLLKEQMRQQKLLVAGGVP
HDACX_V2    LRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALLQHLLLKEQMRQQKLLVAGGVP
```

FIG. 25A

```
BMY_HDAL1   ------------------------------------------------------------
BMY_HDAL2   ------------------------------------------------------------
BMY_HDAL3   ------------------------------------------------------------
HDAC9C      LHPQSPLATKERISPGIRGTHKLPRHRPLNRTQSAPLPQSTLAQLVIQQQHQQFLEKQKQ
HDACX_V1    LHPQSPLATKERISPGIRGTHKLPRHRPLNRTQSAPLPQSTLAQLVIQQQHQQFLEKQKQ
HDACX_V2    LHPQSPLATKERISPGIRGTHKLPRHRPLNRTQSAPLPQSTLAQLVIQQQHQQFLEKQKQ

BMY_HDAL1   ------------------------------------------------------------
BMY_HDAL2   ------------------------------------------------------------
BMY_HDAL3   ------------------------------------------------------------
HDAC9C      YQQQIHMNKLLSKSIEQLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDSSACVDD
HDACX_V1    YQQQIHMNKLLSKSIEQLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDSSACVDD
HDACX_V2    YQQQIHMNKELPMTP---------------------------------------------

BMY_HDAL1   ------------------------------------------------------------
BMY_HDAL2   ------------------------------------------------------------
BMY_HDAL3   ------------------------------------------------------------
HDAC9C      TLGQVGAVKVKEEPVDSDEDAQIQEMESGEQAAFMQQPFLEPTHTRALSVRQAPLAAVGM
HDACX_V1    TLGQVGAVKVKEEPVDSDEDAQIQEMESGEQAAFMQQPFLEPTHTRALSVRQAPLAAVGM
HDACX_V2    ------------------------------------------------------------

BMY_HDAL1   ------------------------------------GIAYDPLMLKHQCVCGNSTTHPEH
BMY_HDAL2   ------------------------------------------------------------
BMY_HDAL3   ------------------------------------------------------------
HDAC9C      DGLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGSATGIAYDPLMLKHQCVCGNSTTHPEH
HDACX_V1    DGLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGSATGIAYDPLMLKHQCVCGNSTTHPEH
HDACX_V2    ------------------------------------------------------------

BMY_HDAL1   AGRIQSIWSRLQETGLLNKCERIQGRKASLEEIQLVHSEHHSLLYGTNPLDGQKLDPRIL
BMY_HDAL2   ------------------------------------------------------------
BMY_HDAL3   ------------------------------------------------------------
HDAC9C      AGRIQSIWSRLQETGLLNKCERIQGRKASLEEIQLVHSEHHSLLYGTNPLDGQKLDPRIL
HDACX_V1    AGRIQSIWSRLQETGLLNKCERIQGRKASLKEIQLVHSEHHSLLYGTNPLDGQKLDPRIL
HDACX_V2    ------------------------------------------------------------

BMY_HDAL1   LGDDSQKFFSSLPCGGLGVST---------------------------------------
BMY_HDAL2   ------------------VDSDTIWNELHSSGAARMAVGCVIELASKVASGELKNGFAVV
BMY_HDAL3   ------------------------------------------------------------
HDAC9C      LGDDSQKFFSSLPCGGLGVDSDTIWNELHSSGAARMAVGCVIELASKVASGELKNGFAVV
HDACX_V1    LGDDSQKFFSSLPCGGLGVDSDTIWNELHSSGAARMAVGCVIELASKVASGELKNGFAVV
HDACX_V2    ------------------------------------------------------------

BMY_HDAL1   ------------------------------------------------------------
BMY_HDAL2   RPPGHHAEESTAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNGTQQAFYADPSI
BMY_HDAL3   ------------------------------------------------------------
HDAC9C      RPPGHHAEESTAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNGTQQAFYADPSI
HDACX_V1    RPPGHHAEESTAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNGTQQAFYADPSI
HDACX_V2    ------------------------------------------------------------
```

FIG. 25B

```
BMY_HDAL1      ------------------------------------------------------------
BMY_HDAL2      LYISLHRYDEGNFFPGSGAPNEVGTGLGEGYNINIAWTGGLDPPMGDVEYLEAFRLVLLS
BMY_HDAL3      ------------------------------------------------------RTIVKP
HDAC9C         LYISLHRYDEGNFFPGSGAPNEVGTGLGEGYNINIAWTGGLDPPMGDVEYLEAFRTIVKP
HDACX_V1       LYISLHRYDEGNFFPGSGAPNEVGTGLGEGYNINIAWTGGLDPPMGDVEYLEAFRTIVKP
HDACX_V2       ------------------------------------------------------------

BMY_HDAL1      ------------------------------------------------------------
BMY_HDAL2      L-----------------------------------------------------------
BMY_HDAL3      VAKEFDPDMVLVSAGFDALEGHTPPLGGYKVTAKCFGHLTKQLMTLADGRVVLALEGGHD
HDAC9C         VAKEFDPDMVLVSAGFDALEGHTPPLGGYKVTAKCFGHLTKQLMTLADGRVVLALEGGHD
HDACX_V1       VAKEFDPDMVLVSAGFDALEGHTPPLGGYKVTAKCFGHLTKQLMTLADGRVVLALEGGHD
HDACX_V2       ------------------------------------------------------------

BMY_HDAL1      ------------------------------------------------------------
BMY_HDAL2      ------------------------------------------------------------
BMY_HDAL3      LTAICDASEACVNALLGNELEPLAEDILHQSPNMNAVISLQKIIEIQSKYWKSVRMVAVP
HDAC9C         LTAICDASEACVNALLGNELEPLAEDILHQSPNMNAVISLQKIIEIQSKYWKSVRMVAVP
HDACX_V1       LTAICDASEACVNALLGNELEPLAEDILHQSPNMNAVISLQKIIEIQSKYWKSVRMVAVP
HDACX_V2       ------------------------------------------------------------

BMY_HDAL1      -------------------------------------------------
BMY_HDAL2      -------------------------------------------------
BMY_HDAL3      RGCALAGAQLQEETETVSALASLTVDVEQPFAQEDSRTAGEPMEEEPAL
HDAC9C         RGCALAGAQLQEETETVSALASLTVDVEQPFAQEDSRTAGEPMEEEPAL
HDACX_V1       RGCALAGAQLQEETETVSALASLTVDVEQPFAQEDSRTAGEPMEEEPAL
HDACX_V2       -------------------------------------------------
```

FIG. 25C

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 December 2002 (27.12.2002)

PCT

(10) International Publication Number
WO 2002/102323 A3

(51) International Patent Classification[7]: C12N 15/11, 15/85, 15/86, 1/20, 9/00, 15/63, C07H 21/04, C12Q 1/68, G01N 33/543, 33/577

(21) International Application Number: PCT/US2002/019560

(22) International Filing Date: 14 June 2002 (14.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/298,296 14 June 2001 (14.06.2001) US

(71) Applicant (*for all designated States except US*): **BRISTOL-MYERS SQUIBB COMPANY** [US/US]; P.O BOX 4000, ROUTE 206 and PROVINCELINE ROAD, PRINCETON, NJ 08543-4000 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **JACKSON, Donald, G.** [US/US]; 2641 Main St. Apt. 1, Lawrenceville, NJ 08648 (US). **LORENZI, Matthew, V.** [US/US]; 424 South 7th Street, Philadelphia, PA 19147 (US). **ATTAR, Ricardo, M.** [US/US]; 10 Santina Ct., Lawrenceville, NJ 08648 (US). **GOTTARDIS, Marco** [US/US]; 9 Harris Road, Princeton, NJ 08540 (US).

(74) Agents: **D'AMICO, Stephen** et al.; Bristol-Myers Squibb Company, P.O. Box 4000, Route 206 and Provinceline Road, Princeton, NJ 08543-4000 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report: 31 March 2005

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2002/102323 A3

(54) Title: NOVEL HUMAN HISTONE DEACETYLASES

(57) Abstract: The present invention relates to newly discovered human histone deacetylases (HDACs), also referred to as histone deacetylase-like polypeptides. The polynucleotide sequences and encoded polypeptides of the novel HDACs are encompassed by the invention, as well as vectors comprising these polynucleotides and host cells comprising these vectors. The invention also relates to antibodies that bind to the disclosed HDAC polypeptides, and methods employing these antibodies. Also related are methods of screening for modulators, such as inhibitors or antagonists, or agonists. The invention also relates to diagnostic and therapeutic applications which employ the disclosed HDAC polynucleotides, polypeptides, and antibodies, and HDAC modulators. Such applications can be used with diseases and disorders associated with abnormal cell growth or proliferation, cell differentiation, and cell survival, e.g., neoplastic cell growth, and especially breast and prostate cancers or tumors.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/19560

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 15/11, 15/85, 15/86, 1/20, 9/00, 15/63; C07H 21/04; C12Q 1/68; G01N 33/543, 577
US CL : 536/23.1, 24.5, 24.33; 435/325, 252.1, 193, 320.1, 69.1, 6, 7.1, 7.23; 436/501, 518

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 536/23.1, 24.5, 24.33; 435/325, 252.1, 193, 320.1, 69.1, 6, 7.1, 7.23; 436/501, 518

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WANG et al., HDAC4, a human histone deacetylase related to yeast HDA1, is a transcritional corepressor, Molecular and Cellular Biology, November 1999, vol. 19, pages 7816-7827	1-20
A	ZHOU et al., Cloning and characterization of a histone deacetylase, HDAC9, Proc. Natl. Acad. Sci. USA, 11 September 2001, vol. 98, pages 10572-10577.	1-20

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 January 2005 (18.01.2005)	10 FEB 2005

Name and mailing address of the ISA/US	Authorized officer
Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450	Maria Marlis MISOOK YU, Ph.D.
Facsimile No. (703) 305-3230	Telephone No. 571-272-1600

INTERNATIONAL SEARCH REPORT

Continuation of B. FIELDS SEARCHED Item 3:
Dialog(5, 155), West (USPT, DWPI), sequence databases
Search terms: histone deacetylases, cancer diagnosis, SEQ ID NOs 2, 95, 87, 96, 4, 5, 83.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 December 2002 (27.12.2002)

PCT

(10) International Publication Number
WO 02/102984 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/19051

(22) International Filing Date: 14 June 2002 (14.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/298,173	14 June 2001 (14.06.2001)	US
60/311,686	10 August 2001 (10.08.2001)	US
60/316,995	4 September 2001 (04.09.2001)	US

(71) Applicant (*for all designated States except US*): **SLOAN-KETTERING INSTITUTE FOR CANCER RESEARCH** [US/US]; 1275 York Avenue, New York, NY 10021 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **RICHON, Victoria** [US/US]; 160 Theodore Fremd Street, #A11, Rye, NY 10580 (US). **ZHOU, Xianbo** [CN/US]; 43 Bradley Street, Dobbs Ferry, NY 10522 (US). **RIFKIND, Richard, A.** [US/US]; 425 East 58th Street, #48A, New York, NY 10022 (US). **MARKS, Paul, A.** [US/US]; 7 Rossiter Road, Washington, CT 06793 (US).

(74) Agents: **BROOK, David, E.** et al.; Hamilton, Brook, Smith & Reynolds, P.C., 530 Virginia Road, P.O. Box 9133, Concord, MA 01742-9133 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/102984 A2

(54) Title: HDAC9 POLYPEPTIDES AND POLYNUCLEOTIDES AND USES THEREOF

(57) Abstract: The present invention features substantially pure HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), an HDRP(ΔNLS) polypeptides, and isolated nucleic acid molecules encoding those polypeptides. The present invention also features vectors containing*HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS),* and *HDRP(ΔNLS)* nucleic acid sequences, and cells containing those vectors.

HDAC9 POLYPEPTIDES AND POLYNUCLEOTIDES AND USES THEREOF

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 60/298,173 filed on June 14, 2001, U.S. Provisional Application No. 60/311,686 filed on August 10, 2001, and U.S. Provisional Application No. 60/316,995, filed on September 4, 2001. The entire teachings of the above applications are incorporated herein by reference.

GOVERNMENT SUPPORT

The invention was supported, in whole or in part, by grant CA-0974823 from the National Cancer Institute. The Government has certain rights in the invention.

BACKGROUND OF THE INVENTION

The N-terminal tails of core histones are covalently modified by post-translational modifications, including acetylation and phosphorylation. Evidence suggests that these covalent modifications play important roles in several biological activities involving chromatin, *e.g.*, transcription and replication. Histone deacetylases (HDACs) catalyze the removal of the acetyl group from the lysine residues in the N-terminal tails of nucleosomal core histones resulting in a more compact chromatin structure, a configuration that is generally associated with repression of transcription.

Five proteins and/or open reading frames in yeast (RPD3, HDA1, HOS1, HOS2 and HOS3) that share significant homology in the catalytic domain have been identified as HDACs based upon their sequence homology to human HDAC1. To date, eight HDACs have been identified in mammalian cells, and classified into two classes based on their structure and similarity to yeast RPD3 or HDA1 proteins. Recently, Sir2 family proteins that are structurally unrelated to the five proteins aforementioned have been identified as NAD-dependent HDACs. Class I HDACs are the yeast RPD3 homologs HDAC1, 2, 3, and 8, and are composed primarily of a catalytic domain. Class II HDACs are the yeast HDA1 homologs HDAC4, 5, 6; and

7. HDAC4, 5, and 7 contain a long non-catalytic N-terminal end and a C-terminal HDAC catalytic domain while HDAC6 has two HDAC catalytic domains.

It has also been determined that histone deacetylases can be sensitive to small molecules, including trichostatin A (TSA), trapoxin, and butyrate. For example, the yeast RPD3 and HDA1 and mammalian HDAC1, 2, 3, 4, 5, 6, 7 and 8 are sensitive to inhibition by trichostatin A (TSA). The Sir2 family HDACs, yeast HOS3 and *Drosophila melanogaster* dHDAC6, however, appear to be relatively insensitive to TSA. A class of hybrid bipolar compounds, such as suberoylanilide hydroxamic acid (SAHA) have also been shown to inhibit histone deacetylases and induce terminal differentiation and/or apoptosis in various transformed cells. Examples of such compounds can be found in U.S. Patent Nos. 5,369,108, issued on November 29, 1994, 5,700,811, issued on December 23, 1997, and 5,773,474, issued on June 30, 1998 to Breslow *et al.*, as well as U.S. Patent Nos. 5,055,608, issued on October 8, 1991, and 5,175,191, issued on December 29, 1992 to Marks *et al.*, the entire content of all of which are hereby incorporated by reference.

The identification of the mechanisms by which histones are deacetylated, and the characterization of histone deacetylase function would be of great benefit in understanding how gene transcription is controlled, how the cell cycle is regulated, and how cells are signaled to undergo terminal differentiation and/or apoptosis. Elucidation of such mechanisms can lead to improved therapeutics for many diseases, in particular those characterized by cell proliferation or a lack of cell differentiation or apoptosis, for example, cancer.

SUMMARY OF THE INVENTION

The present invention relates to isolated or recombinant histone deacetylase polypeptides, and isolated histone deacetylase nucleic acid molecules encoding those polypeptides, as well as vectors and cells containing those isolated nucleic acid molecules.

In one aspect of the invention, the isolated or recombinant histone deacetylase polypeptide is selected from a) an isolated or recombinant polypeptide comprising SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; and b) a polypeptide having at least 60% sequence identity with any one

of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. In one embodiment, the isolated or recombinant histone deacetylase polypeptide consists of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. In another embodiment, the isolated or recombinant histone deacetylase polypeptide is mammalian; preferably, the isolated or recombinant histone deacetylase polypeptide is human.

In another aspect, the invention features an isolated nucleic acid molecule selected from a) an isolated nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9; b) a complement of an isolated nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9; c) an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; d) a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; e) a nucleic acid that is hybridizeable under high stringency conditions to a nucleic acid molecule that encodes any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, or SEQ ID NO: 8, or a complement thereof; or f) a nucleic acid molecule that is hybridizeable under high stringency conditions to a nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, or SEQ ID NO: 7; and g) an isolated nucleic acid molecule that has at least 55% sequence identity with any one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, or a complement thereof. In one embodiment, the isolated nucleic acid molecule consists of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9. In another embodiment, the isolated nucleic acid molecule is mammalian; preferably, the isolated nucleic acid molecule is human.

In other aspects, the invention features a vector comprising the isolated histone deacetylase nucleic acid molecule described above, a cell comprising the vector, and a cell comprising the isolated histone deacetylase nucleic acid molecule described above.

In another aspect, the invention features a purified antibody that selectively binds a histone deacetylase polypeptide described above.

In yet another aspect, the invention features a method of identifying a compound that modulates expression of a histone deacetylase nucleic acid molecule described above. The method comprises the steps of a) contacting the nucleic acid molecule with a candidate compound under conditions suitable for expression; and b) assessing the level of expression of the nucleic acid molecule. A candidate compound that increases or decreases expression of the nucleic acid molecule relative to a control is a compound that modulates expression of the nucleic acid molecule. In one embodiment, the method is carried out in a cell or animal. In another embodiment, the method is carried out in a cell free system.

The invention also features a method of treating a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, for example, cancers such as lymphoma, leukemia, melanoma, ovarian cancer, breast cancer, pancreatic cancer, prostate cancer, colon cancer, and lung cancer and myeloproliferative disorders, including polycythemia vera, essential thrombocythemia, agnogenic myeloid metaplasia, and chronic myelogenous leukemia in an individual, comprising administering a compound identified by the above method.

In still another aspect, the invention features a method of identifying a compound that modulates the enzymatic activity of the histone deacetylase polypeptide described above. The method comprises the steps of a) contacting the polypeptide with a candidate compound under conditions suitable for enzymatic reaction; and b) assessing the activity level of the polypeptide. A candidate compound that increases or decreases the activity level of the polypeptide relative to a control is a compound that modulates the enzymatic activity of the polypeptide. In one embodiment, the method is carried out in a cell or animal. In another embodiment, the method is carried out in a cell free system.

In yet another embodiment, the polypeptide is further contacted with a substrate for the polypeptide, wherein the substrate is selected from the group consisting of a cell proliferation disease binding agent, an apoptotic disease binding agent, and a cell differentiation disease binding agent. In one embodiment, the candidate compound is an inhibitor. In another embodiment, candidate compound is an activator.

In another aspect, the invention features a method of identifying a compound that modulates the transcriptional repression activity of the histone deacetylase polypeptide described above. The method comprises the steps of a) contacting the polypeptide with a candidate compound under conditions suitable for a transcriptional repression reaction; and b) assessing the transcriptional repression activity level of the polypeptide. A candidate compound that increases or decreases the transcriptional repression activity level of the polypeptide relative to a control is a compound that modulates the transcriptional repression activity of the polypeptide. In one embodiment, the method is carried out in a cell or animal. In another embodiment, the method is carried out in a cell free system.

In yet another embodiment, the polypeptide is further contacted with a substrate for the polypeptide, wherein the substrate is selected from the group consisting of a cell proliferation disease binding agent, an apoptotic disease binding agent, and a cell differentiation disease binding agent. In one embodiment, the candidate compound is an inhibitor. In another embodiment, candidate compound is an activator.

In another aspect, the invention features a method of identifying a compound that modulates expression of a histone deacetylase nucleic acid molecule described above. The method comprises the steps of a) providing a nucleic acid molecule comprising a promoter region of the histone deacetylase nucleic acid molecule described above, or part of such a promoter region, operably linked to a reporter gene; b) contacting the nucleic acid molecule or with a candidate compound; and c) assessing the level of the reporter gene. A candidate compound that increases or decreases expression of the reporter gene relative to a control is a compound that modulates expression of the histone deacetylase nucleic acid molecule described above. In one embodiment, the method is carried out in a cell.

In still another aspect, the invention features a method of identifying a polypeptide that interacts with a histone deacetylase polypeptide described above in a yeast two-hybrid system. The method comprises the steps of a) providing a first nucleic acid vector comprising a nucleic acid molecule encoding a DNA binding domain and the histone deacetylase polypeptide described above; b) providing a second nucleic acid vector comprising a nucleic acid encoding a transcription

activation domain and a nucleic acid encoding a test polypeptide; c) contacting the first nucleic acid vector with the second nucleic acid vector in a yeast two-hybrid system; and d) assessing transcriptional activation in the yeast two-hybrid system. An increase in transcriptional activation relative to a control indicates that the test polypeptide is a polypeptide that interacts with the histone deacetylase polypeptide described above.

The invention also features a pharmaceutical composition comprising a histone deacetylase polypeptide described above.

In addition, the present invention features a method of diagnosing a cell proliferation disease, an apoptotic disease, or a cell differentiation disease in a subject. The method comprises the steps of a) obtaining a sample from the subject; and b) assessing the level of activity or expression of the histone deacetylase polypeptide described above or the level of the nucleic acid molecule described above in the sample. If the level is increased relative to a control, then the subject has an increased likelihood of having a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, and if the level is decreased relative to a control, then the subject has a decreased likelihood of having a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. In one embodiment, the polypeptide level is assayed using immunohistochemistry techniques. In another embodiment, the nucleic acid molecule level is assayed using *in situ* hybridization techniques.

Compounds and/or polypeptides identified in the above-described screening methods are also part of the present invention.

DESCRIPTION OF THE FIGURES

FIG. 1 is a schematic representation of the order in which FIGS. 1A-1O should be viewed.

FIGS. 1A-1C show the cDNA sequence of *HDAC9* (SEQ ID NO: 1). The arrows and numbers in the *HDAC9* sequence indicate exons. The boxed portion of the sequence indicates the HDAC domain.

FIGS. 1D-1G show the cDNA sequence of *HDAC9a* (SEQ ID NO: 3). The arrows and numbers in the *HDAC9a* sequence indicate exons. The boxed portion of the sequence indicates the HDAC domain.

FIGS. 1H-1I show the cDNA sequence of *HDRP(ΔNLS)* (SEQ ID NO:9).

FIGS. 1J-1L show the cDNA sequence of *HDAC9(ΔNLS)* (SEQ ID NO:5).

FIGS. 1M-1O show the cDNA sequence of *HDAC9a(ΔNLS)* (SEQ ID NO:7).

FIG. 2 is a schematic representation of the order in which FIGS. 2A-2E should be viewed.

FIG. 2A shows the amino acid sequence of HDAC9 (SEQ ID NO: 2).

FIG. 2B shows the amino acid sequence of HDAC9a (SEQ ID NO: 4).

FIG. 2C shows the amino acid sequence of HDAC9(ΔNLS) (SEQ ID NO: 6).

FIG. 2D shows the amino acid sequence of HDAC9a(ΔNLS) (SEQ ID NO: 8).

FIG. 2E shows the amino acid sequence of and HDRP(ΔNLS) (SEQ ID NO: 10).

FIG. 3 is a schematic representation of the order in which FIGS. 3A-3C should be viewed.

FIGS. 3A-3C show an amino acid sequence alignment of HDRP (SEQ ID NO: 11), HDAC9 (SEQ ID NO: 2), HDAC9a (SEQ ID NO: 4), and HDAC4 (SEQ ID NO: 12) polypeptides. Amino acid sequences of HDAC9 (GenBank Accession: AY032737; SEQ ID NO: 2) and HDAC9a (GenBank Accession:AY032738; SEQ ID NO: 4) are aligned with HDRP (GenBank Accession: BAA34464; SEQ D NO: 11) and HDAC4 (GenBank Accession: NP_006028; SEQ ID NO: 12). The identical residues in all proteins are boxed with solid lines. The similar residues are boxed with dotted lines.

FIG. 4 shows a schematic representation of the human *HDAC9* gene structure. The striped boxes represent exons present in isoforms HDRP, HDAC9a, and HDAC9. The lines represent introns. Broken lines are used for larger introns (with size in base pair on top). The 5' untranslated region cDNA and coding region cDNA are represented here. Exons 1-12 encode a non-catalytic domain of the

polypeptides, and exons 14-21 encode the histone deacetylase catalytic domain of the polypeptides, which provide the polypeptides with deacetylase activity.

FIG. 5 is a schematic representation of the order in which FIGS. 5A-5D should be viewed.

FIGS. 5A-5D show the nucleic acid sequence of *HDAC9*, containing all exons expressed in the various isoforms of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* of the present invention (SEQ ID NO:13).

FIG. 6A is a scanned imaged of a multiple human tissue Northern blot that was probed to determine mRNA expression of *HDAC9* using a cDNA probe that recognizes both *HDAC9* and *HDAC9a*. The tissues examined are lane 1, heart; lane 2, brain; lane 3, placenta; lane 4, lung; lane 5, liver; lane 6, skeletal muscle; lane 7, kidney; and lane 8, pancreas. Positions of the RNA size marker in kilobases (kb) are indicated to the left of the blot.

FIG. 6B is a scanned image of an electrophoretic gel showing the results of RT-PCR analyses of mRNA from the same tissues as examined in the Northern blot of FIG. 6A to determine the distribution of *HDAC9* and *HDAC9a* mRNA among these tissues. PCR products were resolved by agarose gel electrophoresis and visualized by ethidium bromide under UV light. A 1-kb DNA ladder was run on both sides of the gel with the size (in kb) indicated on the left. On the right side, the expected products for *HDAC9* and *HDAC9a* are indicated as 9 and 9a, respectively.

FIG. 7 is a graph of HDAC enzymatic activity of HDAC anti-FLAG-immunoprecipitated proteins isolated from vector control, HDAC9-FLAG, and HDAC9a-FLAG transfected 293T cells, as measured in fluorescence units using *FLUOR DE LYS*™ as a substrate in the presence or absence of 1 μM TSA. Results are shown as the mean of three independent assays. The inset is a scanned image of an anti-FLAG Western blot showing the amount of proteins used in the assay. V, Vector control; 9, HDAC9-FLAG; and 9a, HDAC9a-FLAG.

FIG. 8 is a graph of HDAC enzymatic activity of HDAC anti-FLAG-immunoprecipitated proteins isolated from vector control, and HDAC9a-FLAG (treated with 2 μM SAHA or left untreated) transfected 293T cells, as measured by ^{3}H-acetic acid released from ^{3}H-histones in the presence or absence of 2 μM SAHA.

Vector control; HDAC9a, HDAC9a-FLAG; and HDAC9a+, HDAC9a-FLAG + SAHA.

FIG. 9A shows a scanned image of a Western blot of 293T whole cell lysate and anti-FLAG immunoprecipitates from 293T cells transfected with vector, HDAC9-FLAG or HDAC9a-FLAG using antibodies against MEF2 and FLAG. Top panel, anti-MEF2 Western; bottom panel, anti-FLAG Western. L, 293T whole cell lysate; V, vector control IP; 9, HDAC9-FLAG IP; 9a, HDAC9a-FLAG IP.

FIG. 9B is a graph showing the transcription level of p3XMEF2-*Luc* in the presence or absence of pcDNA3 empty vector (-), pCMV-MEF2C, and/or a vector encoding pFLAG-HDAC9 or pFLAG-HDAC9a. p3XMEF2-*Luc* (100 ng) and pRL-TK (5 ng) were transfected into 293T cells with pcDNA3 empty vector (-) or with pCMV-MEF2C (100 ng) (+) along with the indicated amount of pFLAG-HDAC9 or pFLAG-HDAC9a. pFLAG empty vector was used to adjust the DNA to an equal amount in each transfection. The firefly luciferase activity was first normalized to the co-transfected Renilla luciferase activity and the value for MEF2C alone was then set as 1. Results are shown as the mean of three independent transfections +/- standard deviation.

FIG. 10 shows a schematic representation of the HDAC domains of human non-Sir2 family HDACs and HDRP. The boxes represent histone deacetylase (HDAC) domains.

FIG. 11 is a schematic representation of the order in which FIGS. 11A-11F should be viewed.

FIGS. 11A-11F show the nucleotide sequence of the vector pFLAG-CMV-5b-HDAC9 (VR1) (SEQ ID NO: 14). Lowercase letters are vector backbone, uppercase letters are HDAC9 sequence. "Acc" was added at the beginning of the HDAC9 sequence for translation initiation.

FIG. 12 is a schematic representation of the order in which FIGS. 12-1 through 12-66 should be viewed.

FIGS. 12-1 through 12-66 show the nucleotide sequence of the vector pFLAG-CMV-5b-HDAC9a (VR2), with restriction enzyme sites indicated (SEQ ID NO: 14).

FIG. 13 is a schematic representation of the order in which FIGS. 13A-13E should be viewed.

FIGS. 13A-13E show the nucleotide sequence of the vector pFLAG-CMV-5b-HDAC9a (VR2) (SEQ ID NO: 15). Lowercase letters are vector backbone, uppercase letters are HDAC9a sequence. "Acc" was added at the beginning of the HDAC9a sequence for translation initiation.

FIG. 14 is a schematic representation of the order in which FIGS. 14-1 through 14-61 should be viewed.

FIGS. 14-1 through 14-61 show the nucleotide sequence of the vector pFLAG-CMV-5b-HDAC9a (VR2), with restriction enzyme sites indicated (SEQ ID NO: 15).

DETAILED DESCRIPTION OF THE INVENTION

A protein designated HDRP (See Zhou *et al.*, Proc. Natl. Acad. Sci. USA, 97:1056-1061 (2000)) (also called MITR (See Sparrow *et al.*, EMBO J. 18:5085-5098(1999); Zhang *et al.*, J. Biol. Chem., 276:35-39 (2001); and Zhang *et al.*, Proc. Natl. Acad. Sci. USA, 98:7354-7359 (2001)) that is 50% identical to the N-terminal domains of histone deacetylase 4 (HDAC4) and histone deacetylase 5 (HDAC5) was recently identified. The cloning and characterization of a novel histone deacetylase, *HDAC9*, of which HDRP is an alternatively spliced isoform is described herein. The cDNA sequence of *HDAC9* is shown in FIGS. 1A-1C (SEQ ID NO: 1), and the HDAC9 amino acid sequence is shown in FIG. 2A (SEQ ID NO: 2). In addition to cloning *HDAC9*, other alternatively spliced isoforms of HDAC9, designated as HDAC9a (a polypeptide that is 132 amino acids shorter at the C-terminal end than HDAC9), and isoforms of HDAC9, HDAC9a, and HDRP polypeptides that lack the nuclear localization signal (NLS) in the N-terminal non-catalytic end of HDAC9, termed HDAC9(ΔNLS), HDAC9a(ΔNLS), and HDRP(ΔNLS), respectively were also identified. The cDNA sequence of *HDAC9a* is shown in FIGS. 1D-1G (SEQ ID NO: 3), and the HDAC9a amino acid sequence is shown in FIG. 2B (SEQ ID NO: 4). The cDNA sequence of *HDAC9* lacking amino acids encoding an NLS (*HDAC9(ΔNLS)*) is shown in FIGS. 1J-1L (SEQ ID NO: 5), and the HDAC9 lacking an NLS amino acid sequence is shown in FIG. 2C (SEQ ID NO: 6). The cDNA

sequence of *HDAC9a* encoding a polypeptide lacking an NLS (*HDAC9a(ΔNLS)*) is shown in FIGS. 1M-1O (SEQ ID NO: 7), and the HDAC9a lacking an NLS amino acid sequence is shown in FIG. 2D (SEQ ID NO: 8). The cDNA sequence of *HDRP* encoding a polypeptide lacking an NLS (*HDRP(ΔNLS)*) is shown in FIGS. 1H-1I (SEQ ID NO: 9), and the HDRP lacking an NLS amino acid sequence is shown in FIG. 2E (SEQ ID NO: 10).

POLYPEPTIDES OF THE INVENTION

The present invention features isolated or recombinant HDAC9 polypeptides, HDAC9a polypeptides, HDAC9(ΔNLS) polypeptides, HDAC9a(ΔNLS) polypeptides, and HDRP(ΔNLS) polypeptides, and fragments, derivatives, and variants thereof, as well as polypeptides encoded by nucleotide sequences described herein (*e.g.*, other variants). As used herein, the term "polypeptide" refers to a polymer of amino acids, and not to a specific length; thus, peptides, oligopeptides, and proteins are included within the definition of a polypeptide.

As used herein, a polypeptide is said to be "isolated," "substantially pure," or "substantially pure and isolated" when it is substantially free of cellular material, when it is isolated from recombinant or non-recombinant cells, or free of chemical precursors or other chemicals when it is chemically synthesized. Typically, the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide is isolated, substantially pure, or substantially pure and isolated when it has a relative increased concentration or activity of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS), in comparison to total HDAC concentration or activity. Preferably the increased activity or concentration of the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) is at least 2-fold, more preferably, at least 5-fold, and most preferably, at least 10 fold, in comparison to total HDAC concentration or activity. In addition, a polypeptide can be joined to another polypeptide with which it is not normally associated in a cell (*e.g.*, in a "fusion protein") and still be "isolated," "substantially pure," or "substantially pure and isolated." An isolated, substantially pure, or substantially pure and isolated polypeptide may be obtained, for example, using affinity

purification techniques described herein, as well as other techniques described herein and known to those skilled in the art.

By a "histone deacetylase polypeptide" is meant a polypeptide having histone deacetylase activity, transcription repression activity, and/or the ability to deacetylate other substrates, for example, transcription factors, including p53, CoRest, E2F, GATA-1, TFIIe, and TFIIF that normally have a nuclear or cytoplasmic location in a cell. A histone deacetylase polypeptide is also a polypeptide whose activity can be inhibited by molecules having HDAC inhibitory activity. These molecules fall into four general classes: 1) short-chain fatty acids (e.g., 4-phenylbutyrate and valproic acid); 2) hydroxamic acids(e.g. SAHA, Pyroxamide, trichostatin A (TSA), oxamflatin and CHAPs, such as, CHAP1 and CHAP 31); 3) cyclic tetrapeptides (Trapoxin A, Apicidin and Depsipeptide (FK-228, also known as FR9011228); 4) benzamides (e.g., MS-275); and other compounds such as Scriptaid. Examples of such compounds can be found in U.S. Patent Nos. 5,369,108, issued on November 29, 1994, 5,700,811, issued on December 23, 1997, and 5,773,474, issued on June 30, 1998 to Breslow *et al.*, U.S. Patent Nos. 5,055,608, issued on October 8, 1991, and 5,175,191, issued on December 29, 1992 to Marks *et al.*, as well as, Yoshida *et al.*, Bioessays 17, 423-430 (1995), Saito *et al.*, PNAS USA 96, 4592-4597, (1999), Furamai *et al.*, PNAS USA 98 (1), 87-92 (2001), Komatsu *et al.*, Cancer Res. 61(11), 4459-4466 (2001), Su *et al.*, Cancer Res. 60, 3137-3142 (2000), Lee *et al.*, Cancer Res. 61(3), 931-934 and Suzuki *et al.* J. Med. Chem. 42(15), 3001-3003 (1999) the entire content of all of which are hereby incorporated by reference. Examples of such histone deacetylase polypeptides include HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), HDRP(ΔNLS); a substantially pure polypeptide comprising SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; and a polypeptide having preferably at least 60%, more preferably, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% sequence identity to any one of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, as determined using the BLAST program and parameters described herein.

In one embodiment, the histone deacetylase polypeptide has histone deacetylase activity, transcription repression activity, the ability to deacetylate substrates, or is inhibited by trichostatin A or a hybrid polar compound such as

SAHA. In another embodiment, the HDAC9(ΔNLS) polypeptide has any two of the above biological activities. In still another embodiment, the HDAC9(ΔNLS) polypeptide has any three of the above biological activities. In yet another embodiment, the HDAC9(ΔNLS) polypeptide has all of the above biological activities.

An HDAC9 polypeptide is a histone deacetylase polypeptide as described above. An HDAC9 polypeptide preferably has at least 60%, more preferably, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% sequence identity to SEQ ID NO: 2, as determined using the BLAST program and parameters described herein. An HDAC9 polypeptide is also a polypeptide that comprises the amino acids encoded by exons 23, 24, 25 and/or 26, and that does not comprise the amino acids encoded by exon 13 of the *HDAC9* nucleic acid sequence, as shown in FIGS. 1A-1C, FIG. 4, and FIGS. 5A-5D. Preferably, an HDAC9 polypeptide comprises the sequence of SEQ ID NO: 2. More preferably, an HDAC9 polypeptide consists of the sequence of SEQ ID NO: 2. An HDAC polypeptide is also a polypeptide comprising the amino acid sequence of the polypeptide encoded by the nucleic acid sequence of SEQ ID NO: 1.

An HDAC9a polypeptide is a histone deacetylase polypeptide as described above. An HDAC9a polypeptide preferably has at least 60%, more preferably, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% sequence identity to SEQ ID NO: 4, as determined using the BLAST program and parameters described herein. An HDAC9a polypeptide is also a polypeptide that comprises the amino acids encoded by exon 22, and that does not comprise the amino acids encoded by exons 13, 23, 24, 25, or 26 of the *HDAC9* nucleic acid sequence, as shown in FIGS. 1D-1G, FIG. 4, and FIGS. 5A-5D. Preferably, an HDAC9a polypeptide comprises the sequence of SEQ ID NO: 4. More preferably, an HDAC9a polypeptide consists of the sequence of SEQ ID NO: 4. An HDAC9a polypeptide is also a polypeptide comprising the amino acid sequence of the polypeptide encoded by the nucleic acid sequence of SEQ ID NO: 3.

An HDAC9(ΔNLS) is a histone deacetylase polypeptide as described above. An HDAC9(ΔNLS) polypeptide does not comprise a nuclear localization signal (NLS). An HDAC9(ΔNLS) polypeptide preferably has at least 60%, more

preferably, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% sequence identity to SEQ ID NO: 6, as determined using the BLAST program and parameters described herein. An HDAC9(ΔNLS) polypeptide is also a polypeptide that comprises the amino acids encoded by exons 23, 24, 25, and/or 26, and that does not comprise the amino acids encoded by exons 7 or 13 of the *HDAC9* nucleic acid sequence, as shown in FIGS. 1J-1L, and FIGS. 5A-5D. Preferably, an HDAC9(ΔNLS) polypeptide comprises the sequence of SEQ ID NO: 6. More preferably, an HDAC9(ΔNLS) polypeptide consists of the sequence of SEQ ID NO: 6. An HDAC9(ΔNLS) polypeptide is also a polypeptide comprising the amino acid sequence of the polypeptide encoded by the nucleic acid sequence of SEQ ID NO: 5.

An HDAC9a(ΔNLS) polypeptide is a histone deacetylase polypeptide as described above. An HDAC9a(ΔNLS) does not comprise a nuclear localization signal (NLS). An HDAC9a(ΔNLS) polypeptide preferably has at least 60%, more preferably, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% sequence identity to SEQ ID NO: 8, as determined using the BLAST program and parameters described herein. An HDAC9a(ΔNLS) polypeptide is also a polypeptide that comprises the amino acids encoded by exon 22, and that does not comprise the amino acids encoded by exons 7, 13, 23, 24, 25, or 26 of the *HDAC9* nucleic acid sequence, as shown in FIGS. 1M-1O, and FIGS. 5A-5D. Preferably, an HDAC9a(ΔNLS) polypeptide comprises the sequence of SEQ ID NO: 8. More preferably, an HDAC9a(ΔNLS) polypeptide consists of the sequence of SEQ ID NO: 8. An HDAC9a(ΔNLS) polypeptide is also a polypeptide comprising the amino acid sequence of the polypeptide encoded by the nucleic acid sequence of SEQ ID NO: 7.

An HDRP(ΔNLS) polypeptide is a histone deacetylase polypeptide as described above. An HDRP(ΔNLS) does not comprise a nuclear localization signal (NLS). An HDRP(ΔNLS) polypeptide preferably has at least 60%, more preferably, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% sequence identity to SEQ ID NO: 10, as determined using the BLAST program and parameters described herein. An HDRP(ΔNLS) polypeptide is also a polypeptide that does not comprise the amino acids encoded by exons 7 or 13-26 of the *HDAC9* nucleic acid sequence, as shown in FIGS. 1H-1I and FIGS. 5A-5D. Preferably, an HDRP(ΔNLS) polypeptide comprises the sequence of SEQ ID NO: 10. More preferably, an

HDRP(ΔNLS) polypeptide consists of the sequence of SEQ ID NO: 10. An HDRP(ΔNLS) polypeptide is also a polypeptide comprising the amino acid sequence of the polypeptide encoded by the nucleic acid sequence of SEQ ID NO: 9.

The polypeptides of the invention can be purified to homogeneity. It is understood, however, that preparations in which the polypeptide is not purified to homogeneity are useful. The critical feature is that the preparation allows for the desired function of the polypeptide, even in the presence of considerable amounts of other components. Thus, the invention encompasses various degrees of purity. In one embodiment, the language "substantially free of cellular material" includes preparations of the polypeptide having less than about 30% (by dry weight) other proteins (*i.e.*, contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins.

When a polypeptide is recombinantly produced, it can also be substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, less than about 10%, or less than about 5% of the volume of the polypeptide preparation. The language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

In one embodiment, a polypeptide of the invention comprises an amino acid sequence encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, and complements and portions thereof, (*e.g.*, a complement of any one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9 or a portion of any one of SEQ ID NO: 1 or SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9).

The polypeptides of the invention also encompass fragments and sequence variants. Variants include a substantially homologous polypeptide encoded by the

same genetic locus in an organism, *i.e.*, an allelic variant, as well as other variants. Variants also encompass polypeptides derived from other genetic loci in an organism, but having substantial homology to a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, and complements and portions thereof, or having substantial homology to a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of nucleotide sequences encoding any one of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. Variants also include polypeptides substantially homologous or identical to these polypeptides but derived from another organism, *i.e.*, an ortholog. Variants also include polypeptides that are substantially homologous or identical to these polypeptides that are produced by chemical synthesis. Variants also include polypeptides that are substantially homologous or identical to these polypeptides that are produced by recombinant methods.

As used herein, two polypeptides (or a region of the polypeptides) are substantially homologous or identical when the amino acid sequences are at least about 60-65%, typically at least about 70-75%, more typically at least about 80-85%, and most typically greater than about 90-95% or more homologous or identical. A substantially identical or homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid molecule hybridizing to SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, or a portion thereof, under stringent conditions as more particularly described herein, or will be encoded by a nucleic acid molecule hybridizing to a nucleic acid sequence encoding SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, or portion thereof, under stringent conditions as more particularly described herein.

The percent identity of two nucleotide or amino acid sequences can be determined by aligning the sequences for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first sequence). The nucleotides or amino acids at corresponding positions are then compared, and the percent identity between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = # of identical positions/total # of positions x 100). In

certain embodiments, the length of the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), and HDRP(ΔNLS) amino acid or nucleotide sequence aligned for comparison purposes is at least 30%, preferably, at least 40%, more preferably, at least 60%, and even more preferably, at least 70%, 80%, 90%, or 100% of the length of the reference sequence, for example, those sequences provided in FIGS. 1A-1O and 2A-2E. The actual comparison of the two sequences can be accomplished by well-known methods, for example, using a mathematical algorithm. A preferred, non-limiting example of such a mathematical algorithm is described in Karlin *et al.*, Proc. Natl. Acad. Sci. USA, 90:5873-5877 (1993). Such an algorithm is incorporated into the BLASTN and BLASTX programs (version 2.2) as described in Schaffer *et al.*, Nucleic Acids Res., 29:2994-3005 (2001). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, BLASTN) can be used. See http://www.ncbi.nlm.nih.gov, as available on August 10, 2001. In one embodiment, the database searched is a non-redundant (NR) database, and parameters for sequence comparison can be set at: no filters; Expect value of 10; Word Size of 3; the Matrix is BLOSUM62; and Gap Costs have an Existence of 11 and an Extension of 1.

Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0), which is part of the GCG (Accelrys) sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12 , and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art and include ADVANCE and ADAM as described in Torellis and Robotti, Comput. Appl. Biosci., 10: 3-5 (1994); and FASTA described in Pearson and Lipman, Proc. Natl. Acad. Sci USA, 85: 2444-8 (1988).

In another embodiment, the percent identity between two amino acid sequences can be accomplished using the GAP program in the GCG software package (available at http://www.accelrys.com, as available on August 31, 2001) using either a Blossom 63 matrix or a PAM250 matrix, and a gap weight of 12, 10, 8, 6, or 4 and a length weight of 2, 3, or 4. In yet another embodiment, the percent

identity between two nucleic acid sequences can be accomplished using the GAP program in the GCG software package (available at http://www.cgc.com), using a gap weight of 50 and a length weight of 3.

The invention also encompasses HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9aΔNLS, and HDRP(ΔNLS) polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions performed by an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9aΔNLS, or HDRP(ΔNLS) polypeptide encoded by a nucleic acid molecule of the invention. Similarity is determined by conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, Science 247: 1306-1310 (1990).

A variant polypeptide can differ in amino acid sequence by one or more substitutions, deletions, insertions, inversions, fusions, and truncations or a combination of any of these. Further, variant polypeptides can be fully functional or can lack function in one or more activities, for example, in histone deacetylase activity or transcription repression activity. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree. Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncations or a substitution, insertion, inversion, or deletion in a critical residue or critical region, such critical regions include the HDAC domains, which provide the polypeptide

with deacetylase activity, as shown in the nucleic acid sequences of FIGS. 1A-1G, as well as in the schematic of FIG. 4.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, Science, 244: 1081-1085 (1989)). The latter procedure introduces a single alanine mutation at each of the residues in the molecule (one mutation per molecule). The resulting mutant molecules are then tested for biological activity *in vitro*. Sites that are critical for polypeptide activity can also be determined by structural analysis, such as crystallization, nuclear magnetic resonance, or photoaffinity labeling (See Smith *et al.*, J. Mol. Biol., 224: 899-904 (1992); and de Vos *et al.* Science, 255: 306-312 (1992)).

The invention also includes HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), and HDRP(ΔNLS) polypeptide fragments of the polypeptides of the invention. Fragments can be derived from a polypeptide comprising SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or from a polypeptide encoded by a nucleic acid molecule comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9 or a portion thereof and the complements thereof or other variants. The present invention also encompasses fragments of the variants of the polypeptides described herein. Useful fragments include those that retain one or more of the biological activities of the polypeptide as well as fragments that can be used as an immunogen to generate polypeptide-specific antibodies.

Biologically active fragments (peptides that are, for example, 6, 9, 12, 15, 16, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100, or more amino acids in length) can comprise a domain, segment, or motif, for example, an HDAC domain, that has been identified by analysis of the polypeptide sequence using well-known methods, *e.g.*, signal peptides, extracellular domains, one or more transmembrane segments or loops, ligand binding regions, zinc finger domains, DNA binding domains, acylation sites, glycosylation sites, or phosphorylation sites.

Fragments can be discrete (not fused to other amino acids or polypeptides) or can be within a larger polypeptide. Further, several fragments can be comprised within a single larger polypeptide. In one embodiment a fragment designed for

expression in a host can have heterologous pre- and pro-polypeptide regions fused to the amino terminus of the polypeptide fragment and an additional region fused to the carboxyl terminus of the fragment.

The invention thus provides chimeric or fusion polypeptides. These comprise an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9aΔNLS, or HDRP(ΔNLS) polypeptide of the invention operatively linked to a heterologous protein or polypeptide having an amino acid sequence not substantially homologous to the polypeptide. "Operatively linked" indicates that the polypeptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the polypeptide. In one embodiment, the fusion polypeptide does not affect the function of the polypeptide per se. For example, the fusion polypeptide can be a GST-fusion polypeptide in which the polypeptide sequences are fused to the C-terminus of the GST sequences. Other types of fusion polypeptides include, but are not limited to, enzymatic fusion polypeptides, for example, β-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, and Ig fusions. Such fusion polypeptides, particularly poly-His fusions, can facilitate the purification of recombinant polypeptide. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of a polypeptide can be increased by using a heterologous signal sequence. Therefore, in another embodiment, the fusion polypeptide contains a heterologous signal sequence at its N-terminus.

EP-A 0464 533 discloses fusion proteins comprising various portions of immunoglobulin constant regions. The Fc is useful in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262). In drug discovery, for example, human proteins have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists. (See Bennett *et al.*, Journal of Molecular Recognition, 8: 52-58 (1995) and Johanson *et al.*, The Journal of Biological Chemistry, 270,16: 9459-9471 (1995)). Thus, this invention also encompasses soluble fusion polypeptides containing a polypeptide of the invention and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclass (IgG, IgM, IgA, IgE).

A chimeric or fusion polypeptide can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of nucleic acid fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive nucleic acid fragments that can subsequently be annealed and re-amplified to generate a chimeric nucleic acid sequence (see Ausubel *et al.*, "Current Protocols in Molecular Biology," John Wiley & Sons, (1998), the entire teachings of which are incorporated by reference herein). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST protein). A nucleic acid molecule encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide.

The substantially pure, isolated, or substantially pure and isolated HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9aΔNLS, or HDRP(ΔNLS) polypeptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. In one embodiment, the polypeptide is produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the polypeptide is cloned into an expression vector, the expression vector introduced into a host cell, and the polypeptide expressed in the host cell. The polypeptide can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques.

In general, HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9aΔNLS, and HDRP(ΔNLS) polypeptides of the present invention can be used as a molecular weight marker on SDS-PAGE gels or on molecular sieve gel filtration columns using art-recognized methods. The polypeptides of the present invention can be used to raise antibodies or to elicit an immune response. The polypeptides can also be used as a reagent, *e.g.*, a labeled reagent, in assays to quantitatively determine levels of the polypeptide or a molecule to which it binds (*e.g.*, a receptor or a ligand) in biological fluids. The polypeptides can also be used as markers for cells or tissues

in which the corresponding polypeptide is preferentially expressed, either constitutively, during tissue differentiation, or in a diseased state. The polypeptides can be used to isolate a corresponding binding agent, and to screen for peptide or small molecule antagonists or agonists of the binding interaction. The polypeptides of the present invention can also be used as therapeutic agents.

NUCLEIC ACID MOLECULES OF THE INVENTION

The present invention also features isolated *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleic acid molecules.

By a "histone deacetylase nucleic acid molecule" is meant a nucleic acid molecule that encodes a histone deacetylase polypeptide. Such histone nucleic acids include, for example, the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule described in detail herein; an isolated nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9; a complement of an isolated nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9; an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; a nucleic acid that is hybridizeable under high stringency conditions to a nucleic acid molecule that encodes any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, or SEQ ID NO: 8, or a complement thereof; a nucleic acid molecule that is hybridizeable under high stringency conditions to a nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, or SEQ ID NO: 7; and an isolated nucleic acid molecule that has at least 55%, more preferably, 60%, 65%, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% or 99% sequence identity with any one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, or a complement thereof.

An *HDAC9* nucleic acid molecule is a nucleic acid molecule that encodes an HDAC9 polypeptide. In one embodiment, the *HDAC9* nucleic acid molecule is selected from: a nucleic acid molecule that comprises the nucleic acid sequence of SEQ ID NO: 1; a complement of an isolated nucleic acid comprising SEQ ID NO: 1;

an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2; a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2; a nucleic acid that is hybridizeable under high stringency conditions to a nucleic acid molecule that encodes SEQ ID NO: 2; a nucleic acid molecule that is hybridizeable under high stringency conditions to a nucleic acid comprising SEQ ID NO: 1; and an isolated nucleic acid molecule that has preferably, at least 55%, more preferably, 60%, 65%, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% or 99% sequence identity with SEQ ID NO: 1, as determined using the BLAST program and parameters described herein. In another embodiment, the *HDAC9* nucleic acid molecule consists of the nucleic acid sequence of SEQ ID NO: 1.

An *HDAC9a* nucleic acid molecule is a nucleic acid molecule that encodes an HDAC9a polypeptide. An *HDAC9a* nucleic acid molecule preferably has at least 55%, sequence identity to SEQ ID NO: 3, In one embodiment, the *HDAC9a* nucleic acid molecule is selected from: a nucleic acid molecule that comprises the nucleic acid sequence of SEQ ID NO: 3; a complement of an isolated nucleic acid comprising SEQ ID NO: 3; an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 4; a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 4; a nucleic acid that is hybridizeable under high stringency conditions to a nucleic acid molecule that encodes SEQ ID NO: 4; a nucleic acid molecule that is hybridizeable under high stringency conditions to a nucleic acid comprising SEQ ID NO: 3; and an isolated nucleic acid molecule that has preferably, at least 55%, more preferably, 60%, 65%, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% or 99% sequence identity with SEQ ID NO: 3 or a complement thereof, as determined using the BLAST program and parameters described herein. In another embodiment, the *HDAC9a* nucleic acid molecule consists of the nucleic acid sequence of SEQ ID NO: 3.

An *HDAC9(ΔNLS)* nucleic acid molecule is a nucleic acid molecule that encodes an HDAC9(ΔNLS) polypeptide. In one embodiment, the *HDAC9(ΔNLS)* nucleic acid molecule is selected from: a nucleic acid molecule that comprises the nucleic acid sequence of SEQ ID NO: 5; a complement of an isolated nucleic acid comprising SEQ ID NO: 5; an isolated nucleic acid encoding a histone deacetylase

polypeptide of SEQ ID NO: 6; a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 6; a nucleic acid that is hybridizeable under high stringency conditions to a nucleic acid molecule that encodes SEQ ID NO: 6; a nucleic acid molecule that is hybridizeable under high stringency conditions to a nucleic acid comprising SEQ ID NO: 5; and an isolated nucleic acid molecule that has preferably, at least 55%, more preferably, 60%, 65%, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% or 99% sequence identity with SEQ ID NO: 5 or a complement thereof, as determined using the BLAST program and parameters described herein. In another embodiment, the *HDAC9(ΔNLS)* nucleic acid molecule consists of the nucleic acid sequence of SEQ ID NO: 5.

An *HDAC9a(ΔNLS)* nucleic acid molecule is a nucleic acid molecule that encodes an HDAC9a(ΔNLS) polypeptide. In one embodiment, the *HDAC9a(ΔNLS)* nucleic acid molecule is selected from: a nucleic acid molecule that comprises the nucleic acid sequence of SEQ ID NO: 7; a complement of an isolated nucleic acid comprising SEQ ID NO: 7; an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 8; a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 8; a nucleic acid that is hybridizeable under high stringency conditions to a nucleic acid molecule that encodes SEQ ID NO: 8; a nucleic acid molecule that is hybridizeable under high stringency conditions to a nucleic acid comprising SEQ ID NO: 7; and an isolated nucleic acid molecule that has preferably, at least 55%, more preferably, 60%, 65%, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% or 99% sequence identity with SEQ ID NO: 7 or a complement thereof, as determined using the BLAST program and parameters described herein. In another embodiment, the *HDAC9a(ΔNLS)* nucleic acid molecule consists of the nucleic acid sequence of SEQ ID NO: 7.

An "*HDRP(ΔNLS)* nucleic acid molecule" is a nucleic acid molecule that encodes an HDRP(ΔNLS) polypeptide. In one embodiment, the *HDRP(ΔNLS)* nucleic acid molecule is selected from: a nucleic acid molecule that comprises the nucleic acid sequence of SEQ ID NO: 9; a complement of an isolated nucleic acid comprising SEQ ID NO: 9; an isolated nucleic acid encoding a histone deacetylase

polypeptide of SEQ ID NO: 10; a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 10; and an isolated nucleic acid molecule that has preferably, at least 55%, more preferably, 60%, 65%, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% or 99% sequence identity with SEQ ID NO: 9 or a complement thereof, as determined using the BLAST program and parameters described herein.. In another embodiment, the *HDRP(ΔNLS)* nucleic acid molecule consists of the nucleic acid sequence of SEQ ID NO: 9.

The isolated nucleic acid molecules of the present invention can be RNA, for example, mRNA, or DNA, such as cDNA and genomic DNA. DNA molecules can be double-stranded or single-stranded; single stranded RNA or DNA can be either the coding, or sense, strand or the non-coding, or antisense, strand. The nucleic acid molecule can include all or a portion of the coding sequence of the gene and can further comprise additional non-coding sequences such as introns and non-coding 3' and 5' sequences (including regulatory sequences, for example). Additionally, the nucleic acid molecule can be fused to a marker sequence, for example, a sequence that encodes a polypeptide to assist in isolation or purification of the polypeptide. Such sequences include, but are not limited to, those that encode a glutathione-S-transferase (GST) fusion protein and those that encode a hemagglutinin A (HA) polypeptide marker from influenza.

An “isolated,” “substantially pure,” or “substantially pure and isolated” nucleic acid molecule, as used herein, is one that is separated from nucleic acids that normally flank the gene or nucleotide sequence (as in genomic sequences) and/or has been completely or partially purified from other transcribed sequences (*e.g.*, as in an RNA or cDNA library). For example, an isolated nucleic acid of the invention may be substantially isolated with respect to the complex cellular milieu in which it naturally occurs, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system, or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example, as determined by agarose gel electrophoresis or column chromatography such as

HPLC. Preferably, an isolated nucleic acid molecule comprises at least about 50, 80, or 90% (on a molar basis) of all macromolecular species present.

With regard to genomic DNA, the term "isolated" also can refer to nucleic acid molecules that are separated from the chromosome with which the genomic DNA is naturally associated. For example, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotides that flank the nucleic acid molecule in the genomic DNA of the cell from which the nucleic acid molecule is derived.

The *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated. Thus, recombinant DNA contained in a vector is included in the definition of "isolated" as used herein. Also, isolated nucleic acid molecules include recombinant DNA molecules in heterologous host cells, as well as partially or substantially purified DNA molecules in solution. "Isolated" nucleic acid molecules also encompass *in vivo* and *in vitro* RNA transcripts of the DNA molecules of the present invention. An isolated nucleic acid molecule or nucleotide sequence can include a nucleic acid molecule or nucleotide sequence that is synthesized chemically or by recombinant means. Therefore, recombinant DNA contained in a vector are included in the definition of "isolated" as used herein.

Isolated nucleotide molecules also include recombinant DNA molecules in heterologous organisms, as well as partially or substantially purified DNA molecules in solution. *In vivo* and *in vitro* RNA transcripts of the DNA molecules of the present invention are also encompassed by "isolated" nucleotide sequences. Such isolated nucleotide sequences are useful in the manufacture of the encoded polypeptide, as probes for isolating homologous sequences (*e.g.*, from other mammalian species), for gene mapping (*e.g.*, by *in situ* hybridization with chromosomes), or for detecting expression of the gene in tissue (*e.g.*, human tissue), such as by Northern blot analysis.

The present invention also pertains to variant *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleic acid molecules that are not necessarily found in nature but that encode an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide. Thus, for

example, DNA molecules that comprise a sequence that is different from the naturally-occurring *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleotide sequence but which, due to the degeneracy of the genetic code, encode an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide of the present invention are also the subject of this invention.

The invention also encompasses *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleotide sequences encoding portions (fragments), or encoding variant polypeptides such as analogues or derivatives of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide. Such variants can be naturally-occurring, such as in the case of allelic variation or single nucleotide polymorphisms, or non-naturally-occurring, such as those induced by various mutagens and mutagenic processes. Intended variations include, but are not limited to, addition, deletion, and substitution of one or more nucleotides that can result in conservative or non-conservative amino acid changes, including additions and deletions. Preferably, the *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS)* nucleotide (and/or resultant amino acid) changes are silent or conserved; that is, they do not alter the characteristics or activity of the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide. In one preferred embodiment, the nucleotide sequences are fragments that comprise one or more polymorphic microsatellite markers.

Other alterations of the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecules of the invention can include, for example, labeling, methylation, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoamidates, and carbamates), charged linkages (*e.g.*, phosphorothioates or phosphorodithioates), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine or psoralen), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids). Also included are synthetic molecules that mimic nucleic acid molecules in the ability to bind to a designated sequences via hydrogen bonding and other chemical

interactions. Such molecules include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The invention also pertains to *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleic acid molecules that hybridize under high stringency hybridization conditions, such as for selective hybridization, to a nucleotide sequence described herein (*e.g.*, nucleic acid molecules that specifically hybridize to a nucleotide sequence encoding polypeptides described herein, and, optionally, have an activity of the polypeptide). In one embodiment, the invention includes variants described herein that hybridize under high stringency hybridization conditions (*e.g.*, for selective hybridization) to a nucleotide sequence comprising a nucleotide sequence selected from SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9 and the complement of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ D NO: 7, or SEQ ID NO: 9. In another embodiment, the invention includes variants described herein that hybridize under high stringency hybridization conditions (*e.g.*, for selective hybridization) to a nucleotide sequence encoding an amino acid sequence of SEQ ID NO: 2 (HDAC9), SEQ ID NO: 4 (HDAC9a), SEQ ID NO: 6 (HDAC9(ΔNLS)), SEQ ID NO: 8 (HDAC9a(ΔNLS)), or SEQ ID NO: 10 (HDRP(ΔNLS)). In a preferred embodiment, the variant that hybridizes under high stringency hybridizations encodes a polypeptide that has a biological activity of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide (*e.g.*, histone deacetylase activity or transcription repression activity).

Such nucleic acid molecules can be detected and/or isolated by specific hybridization (*e.g.*, under high stringency conditions). "Specific hybridization," as used herein, refers to the ability of a first nucleic acid to hybridize to a second nucleic acid in a manner such that the first nucleic acid does not hybridize to any nucleic acid other than to the second nucleic acid (*e.g.*, when the first nucleic acid has a higher similarity to the second nucleic acid than to any other nucleic acid in a sample wherein the hybridization is to be performed). "Stringency conditions" for hybridization is a term of art that refers to the incubation and wash conditions, *e.g.*, conditions of temperature and buffer concentration, that permit hybridization of a particular nucleic acid to a second nucleic acid; the first nucleic acid may be

perfectly (*i.e.*, 100%) complementary to the second, or the first and second may share some degree of complementarity that is less than perfect (*e.g.*, 70%, 75%, 85%, 95%). For example, certain high stringency conditions can be used that distinguish perfectly complementary nucleic acids from those of less complementarity. "High stringency conditions," "moderate stringency conditions," and "low stringency conditions" for nucleic acid hybridizations are explained on pages 2.10.1-2.10.16 and pages 6.3.1-6.3.6 in Current Protocols in Molecular Biology (See Ausubel *et al.*, *supra*, the entire teachings of which are incorporated by reference herein). The exact conditions that determine the stringency of hybridization depend not only on ionic strength (*e.g.*, 0.2XSSC or 0.1XSSC), temperature (*e.g.*, room temperature, 42°C or 68°C), and the concentration of destabilizing agents such as formamide or denaturing agents such as SDS, but also on factors such as the length of the nucleic acid sequence, base composition, percent mismatch between hybridizing sequences, and the frequency of occurrence of subsets of that sequence within other non-identical sequences. Thus, equivalent conditions can be determined by varying one or more of these parameters while maintaining a similar degree of identity or similarity between the two nucleic acid molecules. Typically, conditions are used such that sequences at least about 60%, at least about 70%, at least about 80%, at least about 90% or at least about 95% or more identical to each other remain hybridized to one another. By varying hybridization conditions from a level of stringency at which no hybridization occurs to a level at which hybridization is first observed, conditions that will allow a given sequence to hybridize (*e.g.*, selectively) with the most similar sequences in the sample can be determined.

Exemplary conditions are described in Krause and Aaronson, Methods in Enzymology, 200:546-556 (1991). Also, in, Ausubel, *et al.*, *supra*, which describes the determination of washing conditions for moderate or low stringency conditions. Washing is the step in which conditions are usually set so as to determine a minimum level of complementarity of the hybrids. Generally, starting from the lowest temperature at which only homologous hybridization occurs, each °C by which the final wash temperature is reduced (holding SSC concentration constant) allows an increase by 1% in the maximum extent of mismatching among the

sequences that hybridize. Generally, doubling the concentration of SSC results in an increase in Tm of 17°C. Using these guidelines, the washing temperature can be determined empirically for high, moderate, or low stringency, depending on the level of mismatch sought.

For example, a low stringency wash can comprise washing in a solution containing 0.2XSSC/0.1% SDS for 10 minutes at room temperature; a moderate stringency wash can comprise washing in a prewarmed solution (42°C) solution containing 0.2XSSC/0.1% SDS for 15 minutes at 42°C; and a high stringency wash can comprise washing in prewarmed (68°C) solution containing 0.1XSSC/0.1%SDS for 15 minutes at 68°C. Furthermore, washes can be performed repeatedly or sequentially to obtain a desired result as known in the art. Equivalent conditions can be determined by varying one or more of the parameters given as an example, as known in the art, while maintaining a similar degree of identity or similarity between the target nucleic acid molecule and the primer or probe used.

To determine the percent homology or identity of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of one polypeptide or nucleic acid molecule for optimal alignment with the other polypeptide or nucleic acid molecule). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared, as described above.

The present invention also provides isolated *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleic acid molecules that contain a fragment or portion that hybridizes under highly stringent conditions to a nucleotide sequence comprising a nucleotide sequence selected from SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, and the complement of any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9 and also provides isolated nucleic acid molecules that contain a fragment or portion that hybridizes under highly stringent conditions to a nucleotide sequence encoding an amino acid sequence selected from SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, and SEQ ID NO: 10. The nucleic acid fragments of the invention are at least about 15, preferably, at least about 18, 20, 23, or 25 nucleotides, and can be 30, 40, 50, 100, 200 or more nucleotides in length. Longer

fragments, for example, 30 or more nucleotides in length, that encode antigenic polypeptides described herein are particularly useful, such as for the generation of antibodies as described above.

In a related aspect, the *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleic acid fragments of the invention are used as probes or primers in assays such as those described herein. "Probes" or "primers" are oligonucleotides that hybridize in a base-specific manner to a complementary strand of nucleic acid molecules. Such probes and primers include polypeptide nucleic acids, as described in Nielsen *et al.*, Science, 254, 1497-1500 (1991). As also used herein, the term "primer" in particular refers to a single-stranded oligonucleotide that acts as a point of initiation of template-directed DNA synthesis using well-known methods (*e.g.*, PCR, LCR) including, but not limited to those described herein.

Typically, a probe or primer comprises a region of nucleotide sequence that hybridizes to at least about 15, typically about 20-25, and more typically about 40, 50 or 75, consecutive nucleotides of a nucleic acid molecule comprising a contiguous nucleotide sequence selected from: SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, the complement of any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, and a sequence encoding an amino acid sequence of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

In preferred embodiments, a probe or primer comprises 100 or fewer nucleotides, preferably, from 6 to 50 nucleotides, and more preferably, from 12 to 30 nucleotides. In other embodiments, the probe or primer is at least 70% identical to the contiguous nucleotide sequence or to the complement of the contiguous nucleotide sequence, preferably, at least 80% identical, more preferably, at least 90% identical, even more preferably, at least 95% identical, or even capable of selectively hybridizing to the contiguous nucleotide sequence or to the complement of the contiguous nucleotide sequence. Often, the probe or primer further comprises a label, *e.g.*, radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

The nucleic acid molecules of the invention such as those described above can be identified and isolated using standard molecular biology techniques and the sequence information provided in SEQ ID NO: 1, SEQ ID NO; 3, SEQ ID NO: 5,

SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, and /or SEQ ID NO: 10. For example, nucleic acid molecules can be amplified and isolated by the polymerase chain reaction using synthetic oligonucleotide primers designed based on one or more of the nucleic acid sequences provided above and/or the complement of those sequences. Or such nucleic acid molecules may be designed based on nucleotide sequences encoding one or more of the amino acid sequences provided in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. See generally PCR Technology: Principles and Applications for DNA Amplification (ed. H.A. Erlich, Freeman Press, NY, NY, (1992); PCR Protocols: A Guide to Methods and Applications (Eds. Innis *et al.*, Academic Press, San Diego, CA, (1990); Mattila *et al.*, Nucleic Acids Res., 19: 4967 (1991); Eckert *et al.*, PCR Methods and Applications, 1: 17 (1991); PCR (eds. McPherson *et al.*, IRL Press, Oxford)); and U.S. Patent No. 4,683,202. The nucleic acid molecules can be amplified using cDNA, mRNA, or genomic DNA as a template, cloned into an appropriate vector and characterized by DNA sequence analysis.

Other suitable amplification methods include the ligase chain reaction (LCR) (See Wu and Wallace, Genomics, 4:560 (1989), Landegren *et al.*, Science, 241:1077 (1988)), transcription amplification (Kwoh *et al.*, Proc. Natl. Acad. Sci. USA, 86:1173 (1989)), and self-sustained sequence replication (See Guatelli *et al.*, Proc. Nat. Acad. Sci. USA, 87:1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, that produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

The amplified DNA can be radiolabeled and used as a probe for screening a cDNA library derived from human cells, mRNA in zap express, ZIPLOX, or other suitable vector. Corresponding clones can be isolated, DNA can be obtained following *in vivo* excision, and the cloned insert can be sequenced in either or both orientations by art-recognized methods to identify the correct reading frame encoding a polypeptide of the appropriate molecular weight. For example, the direct analysis of the nucleotide sequence of nucleic acid molecules of the present

invention can be accomplished using well-known methods that are commercially available. See, for example, Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (2nd Ed., CSHP, New York (1989)); Zyskind *et al.*, Recombinant DNA Laboratory Manual, (Acad. Press, (1988)). Using these or similar methods, the polypeptide and the DNA encoding the polypeptide can be isolated, sequenced, and further characterized.

Antisense nucleic acid molecules of the invention can be designed using the nucleotide sequences of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9 and/or the complement of any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9 and/or a portion of those sequences, and/or the complement of those portion or sequences, and/or a sequence encoding the amino acid sequence of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, or encoding a portion of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. Such antisense nucleic acid molecules can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid molecule (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Alternatively, the antisense nucleic acid molecule can be produced biologically using an expression vector into which a nucleic acid molecule has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid molecule will be of an antisense orientation to a target nucleic acid of interest).

In general, the isolated *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleic acid sequences of the invention can be used as molecular weight markers on Southern blots, and as chromosome markers that are labeled to map related gene positions. The nucleic acid sequences can also be used to compare with endogenous DNA sequences in patients to identify genetic disorders (*e.g.*, a predisposition for or susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease), and as probes, such as to hybridize and

discover related DNA sequences or to subtract out known sequences from a sample. The nucleic acid molecules of the present invention can also be used as therapeutic agents.

By a "cell proliferation disease" is meant a disease that is caused by or results in undesirably high levels of cell division, undesirably low levels of apoptosis, or both. For example, cancers such as lymphoma, leukemia, melanoma, ovarian cancer, breast cancer, pancreatic cancer, prostate cancer, colon cancer, and lung cancer are all examples of cell proliferation diseases. Myeloproliferative disorders, including polycythemia vera, essential thrombocythemia, agnogenic myeloid metaplasia, and chronic myelogenous leukemia are also cell proliferation diseases.

By a "cell differentiation disease" is meant a disease that is caused by or results in undesirably low levels of cell differentiation, or by undesirably high levels of cell differentiation. For example, cancers such as lymphoma, leukemia, melanoma, ovarian cancer, breast cancer, pancreatic cancer, prostate cancer, colon cancer, and lung cancer are all examples of cell differentiation diseases. Myeloproliferative disorders, including polycythemia vera, essential thrombocythemia, agnogenic myeloid metaplasia, and chronic myelogenious leukemia are also cell differentiation diseases.

By an "apoptotic disease" is meant a condition in which the apoptotic response is abnormal. This may pertain to a cell or a population of cells that does not undergo cell death under appropriate conditions. For example, normally a cell will die upon exposure to apoptotic-triggering agents, such as chemotherapeutic agents, or ionizing radiation. When, however, a subject has an apoptotic disease, for example, cancer, the cell or a population of cells may not undergo cell death in response to contact with apoptotic-triggering agents. In addition, a subject may have an apoptotic disease when the occurrence of cell death is too low, for example, when the number of proliferating cells exceeds the number of cells undergoing cell death, as occurs in cancer when such cells do not properly differentiate.

An apoptotic disease may also be a condition characterized by the occurrence of undesirably high levels of apoptosis. For example, certain neurodegenerative diseases, including but not limited to Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, multiple sclerosis, restenosis, stroke, and ischemic

brain injury are apoptotic diseases in which neuronal cells undergo undesired cell death.

Other diseases for which the polypeptides and nucleic acid molecules of the present invention may be useful for diagnosing and/or treating include, but are not limited to Huntington's disease.

The *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleic acid molecules of the present invention can further be used to derive primers for genetic fingerprinting, to raise anti-polypeptide antibodies using DNA immunization techniques, and as an antigen to raise anti-DNA antibodies or elicit immune responses. Portions or fragments of the nucleotide sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample.

In addition, the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleotide sequences of the invention can be used to identify and express recombinant polypeptides for analysis, characterization, or therapeutic use, or as markers for tissues in which the corresponding polypeptide is expressed, either constitutively, during tissue differentiation, or in diseased states. The nucleic acid sequences can additionally be used as reagents in the screening and/or diagnostic assays described herein, and can also be included as components of kits (*e.g.*, reagent kits) for use in the screening and/or diagnostic assays described herein.

Standard techniques, such as the polymerase chain reaction (PCR) and DNA hybridization, may be used to clone *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* homologs in other species, for example, mammalian homologs. *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* homologs may be readily identified using low-stringency DNA hybridization or low-stringency PCR with human *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* probes or primers. Degenerate primers encoding human HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or

HDRP(ΔNLS) polypeptides may be used to clone *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* homologs by RT-PCR.

Alternatively, additional *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* homologs can be identified by utilizing consensus sequence information for HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptides to search for similar polypeptides in other species. For example, polypeptide databases for other species can be searched for proteins with the HDAC domains described herein. Candidate polypeptides containing such a motif can then be tested for their HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) biological activities, using methods described herein.

EXPRESSION OF THE NUCLEIC ACID MOLECULES OF THE INVENTION

Another aspect of the invention pertains to nucleic acid constructs containing an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule, for example, one selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, and the complement of any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9 (or portions thereof). Yet another aspect of the invention pertains to *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleic acid constructs containing a nucleic acid molecule encoding the amino acid sequence of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. The constructs comprise a vector (*e.g.*, an expression vector) into which a sequence of the invention has been inserted in a sense or antisense orientation.

As used herein, the term "vector" or "construct" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal

mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses) that serve equivalent functions.

Preferred recombinant expression vectors of the invention comprise a nucleic acid molecule of the invention in a form suitable for expression of the nucleic acid molecule in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences).

It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed and the level of expression of polypeptide desired. The expression vectors of the invention can be introduced into host cells to thereby produce polypeptides, including fusion polypeptides, encoded by nucleic acid molecules as described herein.

The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic or eukaryotic cells, *e.g.*, bacterial cells, such as *E. coli*, insect cells (using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example, using T7 promoter regulatory sequences and T7 polymerase.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a nucleic acid molecule of the invention can be expressed in bacterial cells (*e.g.*, *E. coli*), insect cells, yeast, or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells, human 293T cells, HeLa cells, NIH 3T3 cells, and mouse erythroleukemia (MEL) cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing a foreign nucleic acid molecule (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al. (supra)*, and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select

these integrants, a gene that encodes a selectable marker (*e.g.*, for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those that confer resistance to drugs, such as G418, hygromycin, or methotrexate. Nucleic acid molecules encoding a selectable marker can be introduced into a host cell on the same vector as the nucleic acid molecule of the invention or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid molecule can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a polypeptide of the invention. Accordingly, the invention further provides methods for producing a polypeptide using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule of the invention has been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous nucleotide sequences have been introduced into the genome or homologous recombinant animals in which endogenous nucleotide sequences have been altered. Such animals are useful for studying the function and/or activity of the nucleotide sequence and polypeptide encoded by the sequence and for identifying and/or evaluating modulators of their activity.

As used herein, a "transgenic animal" is a non-human animal, preferably, a mammal, more preferably, a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians. A

transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably, a mammal, more preferably, a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191, and in Hogan, Manipulating the Mouse Embryo (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (1986)). Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, Current Opinion in Bio/Technology, 2:823-829 (1991) and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169. Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut *et al.*, Nature, 385:810-813 (1997) and PCT Publication Nos. WO 97/07668 and WO 97/07669.

ANTIBODIES OF THE INVENTION

Polyclonal and/or monoclonal antibodies that selectively bind one form of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide but not another form of the polypeptide are also provided. Antibodies are also provided that bind a portion of either the variant or reference HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide that contains the polymorphic site or sites.

In another aspect, the invention provides antibodies to each of the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), and HDRP(ΔNLS) polypeptides and polypeptide fragments of the invention, *e.g.*, having an amino acid sequence encoded

by SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, or a portion thereof, or having an amino acid sequence encoded by a nucleic acid molecule comprising all or a portion of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9, (*e.g.*, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or another variant, or portion thereof).

The term "purified antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site that selectively binds an antigen. A molecule that selectively binds to a polypeptide of the invention is a molecule that binds to that polypeptide or a fragment thereof, but does not substantially bind other molecules in a sample, *e.g.*, a biological sample that naturally contains the polypeptide. Preferably the antibody is at least 60%, by weight, free from proteins and naturally occurring organic molecules with which it naturally associated. More preferably, the antibody preparation is at least 75% or 90%, and most preferably, 99%, by weight, antibody. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')2 fragments that can be generated by treating the antibody with an enzyme such as pepsin.

The invention provides polyclonal and monoclonal antibodies that selectively bind to an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide of the invention. The term "monoclonal antibody" or "monoclonal antibody composition," as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of a polypeptide of the invention. A monoclonal antibody composition thus typically displays a single binding affinity for a particular polypeptide of the invention with which it immunoreacts.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a desired immunogen, *e.g.*, an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide of the invention or fragment thereof. The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules directed against the polypeptide can be isolated from the mammal (*e.g.*, from the

blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction.

At an appropriate time after immunization, *e.g.*, when the antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein, Nature, 256:495-497 (1975), the human B cell hybridoma technique (Kozbor *et al.*, Immunol. Today, 4:72 (1983)), the EBV-hybridoma technique (Cole *et al.*, Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96 (1985)) or trioma techniques. The technology for producing hybridomas is well known (see generally Current Protocols in Immunology, Coligan *et al.*, (eds.) John Wiley & Sons, Inc., New York, NY (1994)). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with an immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds a polypeptide of the invention.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating a monoclonal antibody to a polypeptide of the invention (see, *e.g.*, Current Protocols in Immunology, *supra*; Galfre *et al.*, (1977) Nature, 266:55052; R.H. Kenneth, in Monoclonal Antibodies: A New Dimension In Biological Analyses, Plenum Publishing Corp., New York, New York (1980); and Lerner, Yale J. Biol. Med., 54:387-402 (1981)). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods that also would be useful.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody to an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with the polypeptide to thereby isolate immunoglobulin library members that bind the polypeptide. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia Recombinant Phage Antibody System, Catalog No. 27-9400-01; and the Stratagene SurfZAP™ Phage

Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs *et al.*, Bio/Technology, 9:1370-1372 (1991); Hay *et al.*, Hum. Antibod. Hybridomas, 3:81-85 (1992); Huse *et al.*, Science, 246:1275-1281 (1989); and Griffiths *et al.*, EMBO J., 12:725-734 (1993).

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art.

In general, antibodies of the invention (*e.g.*, a monoclonal antibody) can be used to isolate an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide of the invention by standard techniques, such as affinity chromatography or immunoprecipitation. A polypeptide-specific antibody can facilitate the purification of natural polypeptide from cells and of recombinantly produced polypeptide expressed in host cells. Moreover, an antibody specific for an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide of the invention can be used to detect the polypeptide (*e.g.*, in a cellular lysate, cell supernatant, or tissue sample) in order to evaluate the abundance and pattern of expression of the polypeptide.

The antibodies of the present invention can also be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, and acetylcholinesterase; examples of suitable

prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride and phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S, and ^{3}H.

DIAGNOSTIC AND SCREENING ASSAYS OF THE INVENTION

The present invention also pertains to diagnostic assays for assessing *HDAC 9 HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* gene expression, or for assessing activity of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptides of the invention. In one embodiment, the assays are used in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, or is at risk for (has a predisposition for or a susceptibility to) developing a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. The invention also provides for prognostic (or predictive) assays for determining whether an individual is susceptible to developing a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. For example, mutations in the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of symptoms associated with a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

Another aspect of the invention pertains to assays for monitoring the influence of agents, or candidate compounds (*e.g.*, drugs or other agents) on the nucleic acid molecule expression or biological activity of polypeptides of the invention, as well as to assays for identifying candidate compounds that bind to an HDAC9, HDAC9a polypeptide, an HDAC9(ΔNLS) polypeptide, an HDAC9a(ΔNLS) polypeptide, or an HDRP(ΔNLS) polypeptide. These and other assays and agents are described in further detail in the following sections.

DIAGNOSTIC ASSAYS

HDAC9, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecules, probes, primers, polypeptides, and antibodies to an HDAC9, an HDAC9a protein, an HDAC9(ΔNLS) protein, an HDAC9a(ΔNLS) protein, or an HDRP(ΔNLS) protein can be used in methods of diagnosis of a susceptibility to, or likelihood of having a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, as well as in kits useful for diagnosis of a susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

In one embodiment of the invention, diagnosis of a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease is made by detecting a polymorphism in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. The polymorphism can be a mutation in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, such as the insertion or deletion of a single nucleotide, or of more than one nucleotide, resulting in a frame shift mutation; the change of at least one nucleotide, resulting in a change in the encoded amino acid; the change of at least one nucleotide, resulting in the generation of a premature stop codon; the deletion of several nucleotides, resulting in a deletion of one or more amino acids encoded by the nucleotides; the insertion of one or several nucleotides, such as by unequal recombination or gene conversion, resulting in an interruption of the coding sequence of the gene; duplication of all or a part of the gene; transposition of all or a part of the gene; or rearrangement of all or a part of the gene, or a change in the expression pattern of the various HDAC9 isoforms. More than one such mutation may be present in a single nucleic acid molecule.

Such sequence changes cause a mutation in the polypeptide encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. For example, if the mutation is a frame shift mutation, the frame shift can result in a change in the encoded amino acids, and/or can result in the generation of a premature stop codon, causing generation of a truncated polypeptide. Alternatively, a polymorphism associated with a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease can be a synonymous

mutation in one or more nucleotides (*i.e.*, a mutation that does not result in a change in the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide). Such a polymorphism may alter sites, affect the stability or transport of mRNA, or otherwise affect the transcription or translation of the nucleic acid molecule. *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS),* or *HDRP(ΔNLS)* that has any of the mutations described above is referred to herein as a "mutant nucleic acid molecule."

In a first method of diagnosing a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, hybridization methods, such as Southern analysis, Northern analysis, or *in situ* hybridizations, can be used (see Ausubel, *et al., supra*). For example, a biological sample from a test subject (a "test sample") of genomic DNA, RNA, or cDNA, is obtained from an individual suspected of having, being susceptible to or predisposed for, or carrying a defect for, a cell proliferation disease, an apoptotic disease, or a cell differentiation disease (the "test individual"). The individual can be an adult, child, or fetus. The test sample can be from any source that contains genomic DNA, such as a blood sample, sample of amniotic fluid, sample of cerebrospinal fluid, or tissue sample from skin, muscle, buccal or conjunctival mucosa, placenta, gastrointestinal tract, or other organs. A test sample of DNA from fetal cells or tissue can be obtained by appropriate methods, such as by amniocentesis or chorionic villus sampling. The DNA, RNA, or cDNA sample is then examined to determine whether a polymorphism in *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS),* or *HDRP(ΔNLS)* is present, and/or to determine which variant(s) encoded by *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS),* or *HDRP(ΔNLS)* is present. The presence of the polymorphism or variant(s) can be indicated by hybridization of the gene in the genomic DNA, RNA, or cDNA to a nucleic acid probe. A "nucleic acid probe," as used herein, can be a DNA probe or an RNA probe; the nucleic acid probe can contain at least one polymorphism in *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS),* or *HDRP(ΔNLS)* or contains a nucleic acid encoding a particular variant of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS). The probe can be any of the nucleic acid

molecules described above (*e.g.*, the entire nucleic acid molecule, a fragment, a vector comprising the gene, a probe, or primer, etc.).

To diagnose a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, a hybridization sample is formed by contacting the test sample containing *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, with at least one nucleic acid probe. A preferred probe for detecting mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* mRNA or genomic DNA sequences described herein. The nucleic acid probe can be, for example, a full-length nucleic acid molecule, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250, or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to appropriate mRNA or genomic DNA. For example, the nucleic acid probe can be all or a portion of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, or the complement of SEQ ID NO: 1 or SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9; or can be a nucleic acid molecule encoding all or a portion of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. Other suitable probes for use in the diagnostic assays of the invention are described above (see. *e.g.*, probes and primers discussed under the heading, "Nucleic Acids of the Invention").

The hybridization sample is maintained under conditions that are sufficient to allow specific hybridization of the nucleic acid probe to *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. "Specific hybridization," as used herein, indicates exact hybridization (*e.g.*, with no mismatches). Specific hybridization can be performed under high stringency conditions or moderate stringency conditions, for example, as described above. In a particularly preferred embodiment, the hybridization conditions for specific hybridization are high stringency.

Specific hybridization, if present, is then detected using standard methods. If specific hybridization occurs between the nucleic acid probe and *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* in the test sample, then *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* has the

polymorphism, or is the variant, that is present in the nucleic acid probe. More than one nucleic acid probe can also be used concurrently in this method. Specific hybridization of any one of the nucleic acid probes is indicative of a polymorphism in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or of the presence of a particular variant encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, and is therefore diagnostic for a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

In Northern analysis (see Current Protocols in Molecular Biology, Ausubel, *et al.*, *supra*), the hybridization methods described above are used to identify the presence of a polymorphism or of a particular variant, associated with a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. For Northern analysis, a test sample of RNA is obtained from the individual by appropriate means. Specific hybridization of a nucleic acid probe, as described above, to RNA from the individual is indicative of a polymorphism in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or of the presence of a particular variant encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, and is therefore diagnostic for a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

For representative examples of use of nucleic acid probes, see, for example, U.S. Patent Nos. 5,288,611 and 4,851,330.

Alternatively, a peptide nucleic acid (PNA) probe can be used instead of a nucleic acid probe in the hybridization methods described above. PNA is a DNA mimic having a peptide-like, inorganic backbone, such as N-(2-aminoethyl)glycine units, with an organic base (A, G, C, T, or U) attached to the glycine nitrogen via a methylene carbonyl linker (see, for example, Nielsen *et al.*, Bioconjugate Chemistry, 5 (1994), American Chemical Society, p. 1 (1994)). The PNA probe can be designed to specifically hybridize to a gene having a polymorphism associated with a susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. Hybridization of the PNA probe to *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* is diagnostic for a decreased

susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

In another method of the invention, mutation analysis by restriction digestion can be used to detect a mutant nucleic acid molecule, or nucleic acid molecules containing a polymorphism(s), if the mutation or polymorphism in the gene results in the creation or elimination of a restriction site. A test sample containing genomic DNA is obtained from the individual. Polymerase chain reaction (PCR) can be used to amplify *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* (and, if necessary, the flanking sequences) in the test sample of genomic DNA from the test individual. RFLP analysis is conducted as described (see Current Protocols in Molecular Biology, *supra*). The digestion pattern of the relevant DNA fragment indicates the presence or absence of the mutation or polymorphism in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, and therefore indicates the presence or absence of this decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

Sequence analysis can also be used to detect specific polymorphisms in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. A test sample of DNA or RNA is obtained from the test individual. PCR or other appropriate methods can be used to amplify the nucleic acid molecule, and/or its flanking sequences, if desired. The sequence of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or HDRP(ΔNLS), or a fragment of the any of those nucleic acid molecules, or an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* cDNA, or a fragment of any of those cDNAs, or an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* mRNA, or a fragment of any of those mRNAs, is determined, using standard methods. The sequence of the above gene, gene fragment, cDNA, cDNA fragment, mRNA, or mRNA fragment is compared with the known nucleic acid sequence of the nucleic acid molecule, cDNA (*e.g.*, SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, or a nucleic acid sequence encoding the protein of SEQ ID NO: 2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO: 10, or a fragment thereof) or mRNA, as appropriate. The presence of a polymorphism in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* indicates that the

individual has a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

Allele-specific oligonucleotides can also be used to detect the presence of a polymorphism in *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, through the use of dot-blot hybridization of amplified oligonucleotides with allele-specific oligonucleotide (ASO) probes (see, for example, Saiki *et al.*, Nature (London) 324:163-166 (1986)). An "allele-specific oligonucleotide" (also referred to herein as an "allele-specific oligonucleotide probe") is an oligonucleotide of approximately 10-50 base pairs, preferably approximately 15-30 base pairs, that specifically hybridizes to *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, and that contains a polymorphism associated with a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. An allele-specific oligonucleotide probe that is specific for particular polymorphisms in *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* can be prepared, using standard methods (see Current Protocols in Molecular Biology, *supra*).

To identify polymorphisms in the gene that are associated with a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease a test sample of DNA is obtained from the individual. PCR can be used to amplify all or a fragment of *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, and its flanking sequences. The DNA containing the amplified *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* (or a fragment of any of those genes) is dot-blotted, using standard methods (see Current Protocols in Molecular Biology, *supra*), and the blot is contacted with the oligonucleotide probe. The presence of specific hybridization of the probe to the amplified *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* is then detected. Specific hybridization of an allele-specific oligonucleotide probe to DNA from the individual is indicative of a polymorphism in *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, and is therefore indicative of a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

In another embodiment, arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual, can be used to identify polymorphisms in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. For example, in one embodiment, an oligonucleotide array can be used. Oligonucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. These oligonucleotide arrays, also described as "GENECHIPS™," have been generally described in the art, for example, U.S. Patent No. 5,143,854 and PCT patent publication Nos. WO 90/15070 and 92/10092. These arrays can generally be produced using mechanical synthesis methods or light directed synthesis methods that incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis methods. See Fodor *et al.*, Science, 251:767-777 (1991), Pirrung *et al.*, U.S. Patent No. 5,143,854; PCT Publication No. WO 90/15070; Fodor *et al.*, PCT Publication No. WO 92/10092, and U.S. Patent No. 5,424,186, the entire teachings of each of which are incorporated by reference herein. Techniques for the synthesis of these arrays using mechanical synthesis methods are described in, *e.g.*, U.S. Patent No. 5,384,261, the entire teachings of which are incorporated by reference herein.

Once an oligonucleotide array is prepared, a nucleic acid of interest is hybridized to the array and scanned for polymorphisms. Hybridization and scanning are generally carried out by methods described herein and also in, *e.g.*, Published PCT Application Nos. WO 92/10092 and WO 95/11995, and U.S. Patent No. 5,424,186, the entire teachings of which are incorporated by reference herein. In brief, a target nucleic acid sequence that includes one or more previously identified polymorphic markers is amplified by well known amplification techniques, *e.g.*, PCR. Typically, this involves the use of primer sequences that are complementary to the two strands of the target sequence both upstream and downstream from the polymorphism. Asymmetric PCR techniques may also be used. Amplified target, generally incorporating a label, is then hybridized with the array under appropriate conditions. Upon completion of hybridization and washing of the array, the array is scanned to determine the position on the array to which the target sequence

hybridizes. The hybridization data obtained from the scan is typically in the form of fluorescence intensities as a function of location on the array.

Although primarily described in terms of a single detection block, *e.g.*, for detection of a single polymorphism, arrays can include multiple detection blocks, and thus be capable of analyzing multiple, specific polymorphisms. In alternate arrangements, it will generally be understood that detection blocks may be grouped within a single array or in multiple, separate arrays so that varying, optimal conditions may be used during the hybridization of the target to the array. For example, it may often be desirable to provide for the detection of those polymorphisms that fall within G-C rich stretches of a genomic sequence, separately from those falling in A-T rich segments. This allows for the separate optimization of hybridization conditions for each situation.

Additional descriptions of the use of oligonucleotide arrays for detection of polymorphisms can be found, for example, in U.S. Patent Nos. 5,858,659 and 5,837,832, the entire teachings of which are incorporated by reference herein.

Other methods of nucleic acid analysis can be used to detect polymorphisms in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* or variants encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. Representative methods include direct manual sequencing (Church and Gilbert Proc. Natl. Acad. Sci. USA 81: 1991-1995, (1988); Sanger *et al.*, Proc. Natl. Acad. Sci. 74: 5463-5467 (1977); Beavis *et al.*, U.S. Patent No. 5,288,644); automated fluorescent sequencing; single-stranded conformation polymorphism assays (SSCP); clamped denaturing gel electrophoresis (CDGE); denaturing gradient gel electrophoresis (DGGE) (Sheffield *et al.*, Proc. Natl. Acad. Sci. USA 86: 232-236 (1991)), mobility shift analysis (Orita *et al.*, Proc. Natl. Acad. Sci. USA 86: 2766-2770 (1989)), restriction enzyme analysis (Flavell *et al.*, Cell 15: 25 (1978); Geever, *et al.*, Proc. Natl. Acad. Sci. USA 78: 5081 (1981)); heteroduplex analysis; chemical mismatch cleavage (CMC) (Cotton *et al.*, Proc. Natl. Acad. Sci. USA 85: 4397-4401 (1985)); RNase protection assays (Myers *et al.*, Science 230: 1242 (1985)); use of polypeptides that recognize nucleotide mismatches, such as *E. coli* mutS protein; and allele-specific PCR.

In another embodiment of the invention, diagnosis of a susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease can also be made by examining the level of an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid, for example, using in situ hybridization techniques known to one skilled in the art, or by examining the level of expression, activity, and/or composition of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, by a variety of methods, including enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, immunohistochemistry, and immunofluorescence. A test sample from an individual is assessed for the presence of an alteration in the level of an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid or in the expression and/or an alteration in composition of the polypeptide encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or for the presence of a particular variant encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. An alteration in expression of a polypeptide encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* can be, for example, an alteration in the quantitative polypeptide expression (*i.e.*, the amount of polypeptide produced); an alteration in the composition of a polypeptide encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or an alteration in the qualitative polypeptide expression (*e.g.*, expression of a mutant HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide or variant thereof). In a preferred embodiment, diagnosis of a susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease is made by detecting a particular variant encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or a particular pattern of variants. Preferably, increased levels of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* or increased expression or activity of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, relative to a control sample, for example, a sample known not to be associated with a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, indicates an increased susceptibility or likelihood that the individual has a cell proliferation disease, an apoptotic disease, or a cell

differentiation disease. Alternatively, decreased levels of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* or decreased expression or activity of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, relative to a control sample, for example, a sample known not to be associated with a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, indicates a decreased susceptibility or likelihood that the individual has a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

Both quantitative and qualitative alterations can also be present. An "alteration" or "modulation" in the polypeptide expression, activity, or composition, as used herein, refers to an alteration in expression or composition in a test sample, as compared with the expression or composition of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide in a control sample. A control sample is a sample that corresponds to the test sample (*e.g.*, is from the same type of cells), and is from an individual who is not affected by a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. An alteration in the expression or composition of the polypeptide in the test sample, as compared with the control sample, is indicative of a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. Similarly, the presence of one or more different variants in the test sample, or the presence of significantly different amounts of different variants in the test sample, as compared with the control sample, is indicative of a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

It is understood that alterations or modulations in polypeptide expression or function can occur in varying degrees. For example, an alteration or modulation in expression can be an increase, for example, by at least 1.5-fold to 2-fold, at least 3-fold, or, at least 5-fold, relative to the control. Alternatively, the alteration or modulation in polypeptide expression can be a decrease, for example, by at least 10%, at least 40%, 50%, or 75%, or by at least 90%, relative to the control.

Various means of examining expression or composition of the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide can be used, including spectroscopy, colorimetry, electrophoresis, isoelectric focusing, and

immunoassays (*e.g.*, David *et al.*, U.S. Patent No. 4,376,110) such as immunoblotting (see also Ausubel *et al.*, *supra*; particularly chapter 10). For example, in one embodiment, an antibody capable of binding to the polypeptide (*e.g.*, as described above), preferably an antibody with a detectable label, can be used. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')2) can be used. The term "labeled," with regard to the antibody, is intended to encompass direct labeling of the antibody by coupling (*i.e.*, physically linking) a detectable substance to the antibody, as well as indirect labeling of the antibody by reacting it with another reagent that is directly labeled. An example of indirect labeling is detection of a primary antibody using a fluorescently labeled secondary antibody.

Western blotting analysis, using an antibody as described above that specifically binds to a mutant HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, or an antibody that specifically binds to a non-mutant HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, or an antibody that specifically binds to a particular variant encoded by *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, can be used to identify the presence in a test sample of a particular variant of a polypeptide encoded by a polymorphic or mutant *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or the absence in a test sample of a particular variant or of a polypeptide encoded by a non-polymorphic or non-mutant gene. The presence of a polypeptide encoded by a polymorphic or mutant gene, or the absence of a polypeptide encoded by a non-polymorphic or non-mutant gene, is diagnostic for a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, as is the presence (or absence) of particular variants encoded by the *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule.

In one embodiment of this method, the level or amount of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide in a test sample is compared with the level or amount of the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide in a control sample. A level or amount of the polypeptide in the test sample that is higher or

lower than the level or amount of the polypeptide in the control sample, such that the difference is statistically significant, is indicative of an alteration in the expression of the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, and is diagnostic for a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

Alternatively, the composition of the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide in a test sample is compared with the composition of the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide in a control sample. A difference in the composition of the polypeptide in the test sample, as compared with the composition of the polypeptide in the control sample (*e.g.*, the presence of different variants), is diagnostic for a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. In another embodiment, both the level or amount and the composition of the polypeptide can be assessed in the test sample and in the control sample. A difference in the amount or level of the polypeptide in the test sample, compared to the control sample; a difference in composition in the test sample, compared to the control sample; or both a difference in the amount or level, and a difference in the composition, is indicative of a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

Kits (*e.g.*, reagent kits) useful in the methods of diagnosis comprise components useful in any of the methods described herein, including, for example, hybridization probes or primers as described herein (*e.g.*, labeled probes or primers), reagents for detection of labeled molecules, restriction enzymes (*e.g.*, for RFLP analysis), allele-specific oligonucleotides, antibodies that bind to a mutant or to non-mutant (native) HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, means for amplification of nucleic acids comprising *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or means for analyzing the nucleic acid sequence of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or for analyzing the amino acid sequence of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, etc.

SCREENING ASSAYS AND AGENTS IDENTIFIED THEREBY

The invention provides methods (also referred to herein as "screening assays") for identifying the presence of a nucleotide that hybridizes to a nucleic acid of the invention, as well as for identifying the presence of a polypeptide encoded by a nucleic acid of the invention. In one embodiment, the presence (or absence) of a nucleic acid molecule of interest (*e.g.*, a nucleic acid that has significant homology with a nucleic acid of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*) in a sample can be assessed by contacting the sample with a nucleic acid comprising a nucleic acid of the invention (*e.g.*, a nucleic acid having the sequence of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9, which may optionally comprise at least one polymorphism, or the complement thereof, or a nucleic acid encoding an amino acid having the sequence of SEQ ID NO: 2, SEQ ID NO:4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or a fragment or variant of such nucleic acids), under stringent conditions as described above, and then assessing the sample for the presence (or absence) of hybridization. In a preferred embodiment, high stringency conditions are conditions appropriate for selective hybridization. In another embodiment, a sample containing the nucleic acid molecule of interest is contacted with a nucleic acid containing a contiguous nucleotide sequence (*e.g.*, a primer or a probe as described above) that is at least partially complementary to a part of the nucleic acid molecule of interest (*e.g.*, an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid), and the contacted sample is assessed for the presence or absence of hybridization. In a preferred embodiment, the nucleic acid containing a contiguous nucleotide sequence is completely complementary to a part of the nucleic acid molecule of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*.

In any of the above embodiments, all or a portion of the nucleic acid of interest can be subjected to amplification prior to performing the hybridization.

In another embodiment, the presence (or absence) of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, such as a polypeptide of the invention or a fragment or variant thereof, in a sample can be assessed by contacting the sample with an antibody that specifically binds to the

polypeptide of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) (*e.g.*, an antibody such as those described above), and then assessing the sample for the presence (or absence) of binding of the antibody to the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide.

In another embodiment, the invention provides methods for identifying agents or compounds (*e.g.*, fusion proteins, polypeptides, peptidomimetics, prodrugs, receptors, binding agents, antibodies, small molecules or other drugs, or ribozymes) that alter or modulate (*e.g.*, increase or decrease) the activity of the polypeptides described herein, or that otherwise interact with the polypeptides herein. For example, such compounds can be compounds or agents that bind to polypeptides described herein (*e.g.*, HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) substrates or agents); that have a stimulatory or inhibitory effect on, for example, activity of polypeptides of the invention; or that change (*e.g.*, enhance or inhibit) the ability of the polypeptides of the invention to interact with HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) binding agents; or that alter post-translational processing of the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide (*e.g.*, agents that alter proteolytic processing to direct the polypeptide from where it is normally synthesized to another location in the cell, such as the cell surface; or agents that alter proteolytic processing such that more polypeptide is released from the cell, etc.). In one example, the binding agent is a cell proliferation disease binding agent, an apoptotic disease binding agent, or a cell differentiation disease binding agent. As used herein, by a "cell proliferation disease binding agent," an "apoptotic disease binding agent," or a "cell differentiation disease binding agent" is meant an agent as described herein that binds to a polypeptide of the present invention and modulates a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. The modulation can be an increase or a decrease in the severity or progression of the disease. In addition, a cell proliferation disease binding agent, an apoptotic disease binding agent, or a cell differentiation disease binding agent includes an agent that binds to a polypeptide that is upstream (earlier) or downstream (later) of the cell signaling events mediated by a polypeptide of the

present invention, and thereby modulates the overall activity of the signaling pathway; in turn, the disease state is modulated.

The candidate compound can cause an increase in the activity of the polypeptide. For example, the activity of the polypeptide can be increased by at least 1.5-fold to 2-fold, at least 3-fold, or, at least 5-fold, relative to the control. Alternatively, the polypeptide activity can be a decrease, for example, by at least 10%, at least 20%, 40%, 50%, or 75%, or by at least 90%, relative to the control.

In one embodiment, the invention provides assays for screening candidate compounds or test agents to identify compounds that bind to or modulate the activity of polypeptides described herein (or biologically active portion(s) thereof), as well as agents identifiable by the assays. As used herein, a "candidate compound" or "test agent" is a chemical molecule, be it naturally-occurring or artificially-derived, and includes, for example, peptides, proteins, synthesized molecules, for example, synthetic organic molecules, naturally-occurring molecule, for example, naturally occurring organic molecules, nucleic acid molecules, and components thereof.

In general, candidate compounds for uses in the present invention may be identified from large libraries of natural products or synthetic (or semi-synthetic) extracts or chemical libraries according to methods known in the art. Those skilled in the field of drug discovery and development will understand that the precise source of test extracts or compounds is not critical to the screening procedure(s) of the invention. Accordingly, virtually any number of chemical extracts or compounds can be screened using the exemplary methods described herein. Examples of such extracts or compounds include, but are not limited to, plant-, fungal-, prokaryotic- or animal-based extracts, fermentation broths, and synthetic compounds, as well as modification of existing compounds. Numerous methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Synthetic compound libraries are commercially available, e.g., from Brandon Associates (Merrimack, NH) and Aldrich Chemical (Milwaukee, WI). Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts are commercially available from a number of sources, including Biotics (Sussex, UK), Xenova

(Slough, UK), Harbor Branch Oceangraphics Institute (Ft. Pierce, FL), and PharmaMar, U.S.A. (Cambridge, MA). In addition, natural and synthetically produced libraries are generated, if desired, according to methods known in the art, e.g., by standard extraction and fractionation methods. For example, candidate compounds can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer or small molecule libraries of compounds (Lam, Anticancer Drug Des., 12: 145 (1997)). Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

In addition, those skilled in the art of drug discovery and development readily understand that methods for dereplication (e.g., taxonomic dereplication, biological dereplication, and chemical dereplication, or any combination thereof) or the elimination of replicates or repeats of materials already known for their activities should be employed whenever possible.

When a crude extract is found to modulate (i.e., stimulate or inhibit) the expression and/or activity of the nucleic acids and or polypeptides of the present invention, further fractionation of the positive lead extract is necessary to isolate chemical constituents responsible for the observed effect. Thus, the goal of the extraction, fractionation, and purification process is the careful characterization and identification of a chemical entity within the crude extract having an activity that stimulates or inhibits nucleic acid expression, polypeptide expression, or polypeptide biological activity. The same assays described herein for the detection of activities in mixtures of compounds can be used to purify the active component and to test derivatives thereof. Methods of fractionation and purification of such heterogenous extracts are known in the art. If desired, compounds shown to be useful agents for treatment are chemically modified according to methods known in the art. Compounds identified as being of therapeutic value may be subsequently analyzed

using animal models for diseases in which it is desirable to alter the activity or expression of the nucleic acids or polypeptides of the present invention.

In one embodiment, to identify candidate compounds that alter the biological activity, for example, the enzymatic activity or transcriptional repression activity of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, a cell, tissue, cell lysate, tissue lysate, or solution containing or expressing an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide (*e.g.*, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SE ID NO: 8, SEQ ID NO: 10, or another variant encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*), or a fragment or derivative thereof (as described above), can be contacted with a candidate compound to be tested under conditions suitable for enzymatic reaction or transcriptional repression reaction, as described herein.

Alternatively, the polypeptide can be contacted directly with the candidate compound to be tested. The level (amount) of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) biological activity is assessed (*e.g.*, the level (amount) of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) biological activity is measured, either directly or indirectly), and is compared with the level of biological activity in a control (*i.e.*, the level of activity of the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide or active fragment or derivative thereof in the absence of the candidate compound to be tested, or in the presence of the candidate compound vehicle only). If the level of the biological activity in the presence of the candidate compound differs, by an amount that is statistically significant, from the level of the biological activity in the absence of the candidate compound, or in the presence of the candidate compound vehicle only, then the candidate compound is a compound that alters the biological activity of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide. For example, an increase in the level of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) enzymatic or transcriptional repression activity relative to a control, indicates that the candidate compound is a compound that enhances (is an agonist of) HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) activity. Similarly,

a decrease in the enzymatic level or transcriptional repression level of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) activity relative to a control, indicates that the candidate compound is a compound that inhibits (is an antagonist of) HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) activity. In another embodiment, the level of biological activity of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide or derivative or fragment thereof in the presence of the candidate compound to be tested, is compared with a control level that has previously been established. A level of the biological activity in the presence of the candidate compound that differs from the control level by an amount that is statistically significant indicates that the compound alters HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) biological activity.

The present invention also relates to an assay for identifying compounds that alter the expression of an *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS)* nucleic acid molecule (*e.g.*, antisense nucleic acids, fusion proteins, polypeptides, peptidomimetics, prodrugs, receptors, binding agents, antibodies, small molecules or other drugs, or ribozymes) that alter (*e.g.*, increase or decrease) expression (*e.g.*, transcription or translation) of the nucleic acid molecule or that otherwise interact with the nucleic acids described herein, as well as compounds identifiable by the assays. For example, a solution containing a nucleic acid encoding an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide can be contacted with a candidate compound to be tested. The solution can comprise, for example, cells containing the nucleic acid or cell lysate containing the nucleic acid; alternatively, the solution can be another solution that comprises elements necessary for transcription/translation of the nucleic acid. Cells not suspended in solution can also be employed, if desired. The level and/or pattern of *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS)* expression (*e.g.*, the level and/or pattern of mRNA or of protein expressed, such as the level and/or pattern of different variants) is assessed, and is compared with the level and/or pattern of expression in a control (*i.e.*, the level and/or pattern of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* expression in the absence of the candidate compound, or in the presence of the candidate

compound vehicle only). If the level and/or pattern in the presence of the candidate compound differs, by an amount or in a manner that is statistically significant, from the level and/or pattern in the absence of the candidate compound, or in the presence of the candidate compound vehicle only, then the candidate compound is a compound that alters the expression of *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. Enhancement of *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* expression indicates that the candidate compound is an agonist of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) activity. Similarly, inhibition of *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* expression indicates that the candidate compound is an antagonist of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) activity. In another embodiment, the level and/or pattern of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide(s) (*e.g.*, different variants) in the presence of the candidate compound to be tested, is compared with a control level and/or pattern that has previously been established. A level and/or pattern in the presence of the candidate compound that differs from the control level and/or pattern by an amount or in a manner that is statistically significant indicates that the candidate compound alters *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* expression.

In another embodiment of the invention, compounds that alter the expression of an *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule or that otherwise interact with the nucleic acids described herein, can be identified using a cell, cell lysate, or solution containing a nucleic acid encoding the promoter region of the *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* gene operably linked to a reporter gene. After contact with a candidate compound to be tested, the level of expression of the reporter gene (*e.g.*, the level of mRNA or of protein expressed) is assessed, and is compared with the level of expression in a control (*i.e.*, the level of the expression of the reporter gene in the absence of the candidate compound, or in the presence of the candidate compound vehicle only). If the level in the presence of the candidate compound differs, by an amount or in a manner that is statistically significant, from

the level in the absence of the candidate compound, or in the presence of the candidate compound vehicle only, then the candidate compound is a compound that alters the expression of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, as indicated by its ability to alter expression of a gene that is operably linked to the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* gene promoter. Enhancement of the expression of the reporter indicates that the compound is an agonist of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) activity. Similarly, inhibition of the expression of the reporter indicates that the compound is an antagonist of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) activity. In another embodiment, the level of expression of the reporter in the presence of the candidate compound to be tested, is compared with a control level that has previously been established. A level in the presence of the candidate compound that differs from the control level by an amount or in a manner that is statistically significant indicates that the candidate compound alters *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* expression.

Compounds that alter the amounts of different variants encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* (*e.g.*, a compound that enhances activity of a first variant, and that inhibits activity of a second variant), as well as compounds that are agonists of activity of a first variant and antagonists of activity of a second variant, can easily be identified using these methods described above.

In other embodiments of the invention, assays can be used to assess the impact of a candidate compound on the activity of a polypeptide in relation to an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) substrate, for example, an inhibitor of histone deacetylase activity. These inhibitors fall into four general classes: 1) short-chain fatty acids (*e.g.*, 4-phenylbutyrate and valproic acid); 2) hydroxamic acids (*e.g.*, SAHA, Pyroxamide, trichostatin A (TSA), oxamflatin and CHAPs, such as, CHAP1 and CHAP 31); 3) cyclic tetrapeptides (Trapoxin A, Apicidin and Depsipeptide (FK-228, also known as FR9011228); 4) benzamides (*e.g.*, MS-275); and other compounds such as Scriptaid. Examples of such assays and compounds can be found in U.S. Patent Nos. 5,369,108, issued on

November 29, 1994, 5,700,811, issued on December 23, 1997, and 5,773,474, issued on June 30, 1998 to Breslow *et al.*, U.S. Patent Nos. 5,055,608, issued on October 8, 1991, and 5,175,191, issued on December 29, 1992 to Marks *et al.*, as well as, Yoshida *et al.*, *supra*; Saito *et al.*, *supra*; Furamai *et al.*, *supra*; Komatsu *et al.*, *supra*; Su *et al.*, *supra*; Lee *et al.*, *supra* and Suzuki *et al. supra*, the entire content of all of which are hereby incorporated by reference.

In one example, a cell or tissue that expresses or contains a compound that interacts with HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) (herein referred to as an "HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) substrate," which can be a polypeptide or other molecule that interacts with HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS)) is contacted with HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) in the presence of a candidate compound, and the ability of the candidate compound to alter the interaction between HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) and the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP (ΔNLS) substrate is determined, for example, by assaying activity of the polypeptide. Alternatively, a cell lysate or a solution containing the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) substrate, can be used. A compound that binds to HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) or the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) substrate can alter the interaction by interfering with, or enhancing the ability of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) to bind to, associate with, or otherwise interact with the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) substrate.

Determining the ability of the candidate compound to bind to HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) or an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) substrate can be accomplished, for example, by coupling the candidate compound with a radioisotope or enzymatic label such that binding of the candidate compound to the polypeptide can be determined by detecting the labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of

radioemmission or by scintillation counting. Alternatively, candidate compound can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a candidate compound to interact with the polypeptide without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a candidate compound with HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) or an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) substrate without the labeling of either the candidate compound, HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS), or the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) substrate (McConnell *et al.*, (1992) Science, 257: 1906-1912). As used herein, a "microphysiometer" (*e.g.*, CYTOSENSOR™) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between ligand and polypeptide.

In another embodiment of the invention, assays can be used to identify polypeptides that interact with one or more HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptides, as described herein. For example, a yeast two-hybrid system such as that described by Fields and Song (Fields and Song, Nature 340: 245-246 (1989)) can be used to identify polypeptides that interact with one or more HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptides. In such a yeast two-hybrid system, vectors are constructed based on the flexibility of a transcription factor that has two functional domains (a DNA binding domain and a transcription activation domain). If the two domains are separated but fused to two different proteins that interact with one another, transcriptional activation can be achieved, and transcription of specific markers (*e.g.*, nutritional markers such as His and Ade, or color markers such as lacZ) can be used to identify the presence of interaction and transcriptional activation. For example, in the methods of the invention, a first vector is used that includes a nucleic acid encoding a DNA binding domain and an HDAC9, HDAC9a,

HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, variant, or fragment or derivative thereof, and a second vector is used that includes a nucleic acid encoding a transcription activation domain and a nucleic acid encoding a polypeptide that potentially may interact with the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, variant, or fragment or derivative thereof (*e.g.*, an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide substrate or receptor). Incubation of yeast containing the first vector and the second vector under appropriate conditions (*e.g.*, mating conditions such as used in the MATCHMAKER™ system from Clontech) allows identification of colonies that express the markers of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS). These colonies can be examined to identify the polypeptide(s) that interact with the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide or fragment or derivative thereof. Such polypeptides may be useful as compounds that alter the activity or expression of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, as described above.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, or an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) substrate, or other components of the assay on a solid support, in order to facilitate separation of complexed from uncomplexed forms of one or both of the polypeptides, as well as to accommodate automation of the assay. Binding of a candidate compound to the polypeptide, or interaction of the polypeptide with a substrate in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein (*e.g.*, a glutathione-S-transferase fusion protein) can be provided that adds a domain that allows HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) or an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) substrate to be bound to a matrix or other solid support.

In another embodiment, modulators of expression of nucleic acid molecules of the invention are identified in a method wherein a cell, cell lysate, tissue, tissue lysate, or solution containing a nucleic acid encoding HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) is contacted with a candidate compound and the expression of appropriate mRNA or polypeptide (*e.g.*, variant(s)) in the cell, cell lysate, tissue, or tissue lysate, or solution, is determined. The level of expression of appropriate mRNA or polypeptide(s) in the presence of the candidate compound is compared to the level of expression of mRNA or polypeptide(s) in the absence of the candidate compound, or in the presence of the candidate compound vehicle only. The candidate compound can then be identified as a modulator of expression based on this comparison. For example, when expression of mRNA or polypeptide is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator or enhancer of the mRNA or polypeptide expression. Alternatively, when expression of the mRNA or polypeptide is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of the mRNA or polypeptide expression. The level of mRNA or polypeptide expression in the cells can be determined by methods described herein for detecting mRNA or polypeptide.

This invention further pertains to novel compounds identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use a compound identified as described herein in an appropriate animal model. For example, a compound identified as described herein (*e.g.*, a candidate compound that is a modulating compound such as an antisense nucleic acid molecule, a specific antibody, or a polypeptide substrate) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such a compound. Alternatively, a compound identified as described herein can be used in an animal model to determine the mechanism of action of such a compound. Furthermore, this invention pertains to uses of novel compounds identified by the above-described screening assays for treatments as described herein. In addition, a compound identified as described herein can be used to alter activity of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, or to

alter expression of *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, by contacting the polypeptide or the nucleic acid molecule (or contacting a cell comprising the polypeptide or the nucleic acid molecule) with the compound identified as described herein.

PHARMACEUTICAL COMPOSITIONS

The present invention also pertains to pharmaceutical compositions comprising nucleic acids described herein, particularly nucleotides encoding the polypeptides described herein; comprising polypeptides described herein (*e.g.*, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO:10, and/or other variants encoded by *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*); and/or comprising a compound that alters (*e.g.*, increases or decreases) *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* expression or HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide activity as described herein. For instance, a polypeptide, protein, fragment, fusion protein or prodrug thereof, or a nucleotide or nucleic acid construct (vector) comprising a nucleotide of the present invention, a compound that alters HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide activity, a compound that alters *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid expression, or an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) substrate or binding partner, can be formulated with a physiologically acceptable carrier or excipient to prepare a pharmaceutical composition. The carrier and composition can be sterile. The formulation should suit the mode of administration.

Suitable pharmaceutically acceptable carriers include but are not limited to water, salt solutions (*e.g.*, NaCl), saline, buffered saline, alcohols, glycerol, ethanol, gum arabic, vegetable oils, benzyl alcohols, polyethylene glycols, gelatin, carbohydrates such as lactose, amylose or starch, dextrose, magnesium stearate, talc, silicic acid, viscous paraffin, perfume oil, fatty acid esters, hydroxymethylcellulose, polyvinyl pyrolidone, etc., as well as combinations thereof. The pharmaceutical preparations can, if desired, be mixed with auxiliary agents, *e.g.*, lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic

pressure, buffers, coloring, flavoring and/or aromatic substances and the like that do not deleteriously react with the active compounds.

The composition, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents. The composition can be a liquid solution, suspension, emulsion, tablet, pill, capsule, sustained release formulation, or powder. The composition can be formulated as a suppository, with traditional binders and carriers such as triglycerides. Oral formulation can include standard carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, polyvinyl pyrollidone, sodium saccharine, cellulose, magnesium carbonate, etc.

Methods of introduction of these compositions include, but are not limited to, intradermal, intramuscular, intraperitoneal, intraocular, intravenous, subcutaneous, topical, oral and intranasal. Other suitable methods of introduction can also include gene therapy (as described below), rechargeable or biodegradable devices, particle acceleration devises ("gene guns") and slow release polymeric devices. The pharmaceutical compositions of this invention can also be administered as part of a combinatorial therapy with other compounds.

The composition can be formulated in accordance with the routine procedures as a pharmaceutical composition adapted for administration to human beings. For example, compositions for intravenous administration typically are solutions in sterile isotonic aqueous buffer. Where necessary, the composition may also include a solubilizing agent and a local anesthetic to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampule or sachette indicating the quantity of active compound. Where the composition is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water, saline or dextrose/water. Where the composition is administered by injection, an ampule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

For topical application, nonsprayable forms, viscous to semi-solid or solid forms comprising a carrier compatible with topical application and having a

dynamic viscosity preferably greater than water, can be employed. Suitable formulations include but are not limited to solutions, suspensions, emulsions, creams, ointments, powders, enemas, lotions, sols, liniments, salves, aerosols, etc., that are, if desired, sterilized or mixed with auxiliary agents, *e.g.*, preservatives, stabilizers, wetting agents, buffers or salts for influencing osmotic pressure, etc. The compound may be incorporated into a cosmetic formulation. For topical application, also suitable are sprayable aerosol preparations wherein the active ingredient, preferably in combination with a solid or liquid inert carrier material, is packaged in a squeeze bottle or in admixture with a pressurized volatile, normally gaseous propellant, *e.g.*, pressurized air.

Compounds described herein can be formulated as neutral or salt forms. Pharmaceutically acceptable salts include those formed with free amino groups such as those derived from hydrochloric, phosphoric, acetic, oxalic, tartaric acids, etc., and those formed with free carboxyl groups such as those derived from sodium, potassium, ammonium, calcium, ferric hydroxides, isopropylamine, triethylamine, 2-ethylamino ethanol, histidine, procaine, etc.

The compounds are administered in a therapeutically effective amount. The amount of compounds that will be therapeutically effective in the treatment of a particular disorder or condition will depend on the nature of the disorder or condition, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the symptoms of a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, and should be decided according to the judgment of a practitioner and each patient's circumstances. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, that notice

reflects approval by the agency of manufacture, use of sale for human administration. The pack or kit can be labeled with information regarding mode of administration, sequence of drug administration (*e.g.*, separately, sequentially or concurrently), or the like. The pack or kit may also include means for reminding the patient to take the therapy. The pack or kit can be a single unit dosage of the combination therapy or it can be a plurality of unit dosages. In particular, the compounds can be separated, mixed together in any combination, present in a single vial or tablet. Compounds assembled in a blister pack or other dispensing means is preferred. For the purpose of this invention, unit dosage is intended to mean a dosage that is dependent on the individual pharmacodynamics of each compound and administered in FDA approved dosages in standard time courses.

METHODS OF THERAPY

The present invention also pertains to methods of treatment (prophylactic, diagnostic, and/or therapeutic) for a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, using an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) therapeutic compound. An "HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) therapeutic compound" is a compound that alters (*e.g.*, enhances or inhibits) HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide activity and/or *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule expression, as described herein (*e.g.*, an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) agonist or antagonist). HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) therapeutic compounds can alter HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide activity or nucleic acid molecule expression by a variety of means, such as, for example, by providing additional HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide or by upregulating the transcription or translation of the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule; by altering post-translational processing of the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide; by altering

transcription of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* variants; or by interfering with HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide activity (*e.g.*, by binding to an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide), or by downregulating the transcription or translation of the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule. Representative HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) therapeutic compounds include the following: nucleic acids or fragments or derivatives thereof described herein, particularly nucleotides encoding the polypeptides described herein and vectors comprising such nucleic acids (*e.g.*, a nucleic acid molecule, cDNA, and/or RNA, such as a nucleic acid encoding an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide or active fragment or derivative thereof, or an oligonucleotide; for example, SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9, which may optionally comprise at least one polymorphism, or a nucleic acid encoding SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, or fragments or derivatives thereof); polypeptides described herein (*e.g.*, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 SEQ ID NO: 10 and/or other variants encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or fragments or derivatives thereof); HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) substrates; peptidomimetics; fusion proteins or prodrugs thereof; antibodies (*e.g.*, an antibody to a mutant HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, or an antibody to a non-mutant HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, or an antibody to a particular variant encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, as described above); ribozymes; other small molecules; and other compounds that alter (*e.g.*, enhance or inhibit) *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid expression or polypeptide activity, for example, those compounds identified in the screening methods described herein, or that regulate transcription of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* variants (*e.g.*,

compounds that affect which variants are expressed, or that affect the amount of each variant that is expressed. More than one HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) therapeutic compound can be used concurrently, if desired.

The HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) therapeutic compound that is a nucleic acid is used in the treatment of a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. The term, "treatment" as used herein, refers not only to ameliorating symptoms associated with the disease, but also preventing or delaying the onset of the disease, and also lessening the severity or frequency of symptoms of the disease. The therapy is designed to alter (*e.g.*, inhibit or enhance), replace or supplement activity of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide in an individual. For example, an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) therapeutic compound can be administered in order to upregulate or increase the expression or availability of the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule or of specific variants of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS), or, conversely, to downregulate or decrease the expression or availability of the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule or specific variants of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS). Upregulation or increasing expression or availability of a native *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule or of a particular variant could interfere with or compensate for the expression or activity of a defective gene or another variant; downregulation or decreasing expression or availability of a native *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule or of a particular variant could minimize the expression or activity of a defective gene or the particular variant and thereby minimize the impact of the defective gene or the particular variant.

The HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) therapeutic compound(s) are administered in a therapeutically effective amount (*i.e.*, an amount that is sufficient to treat the disease, such as by

ameliorating symptoms associated with the disease, preventing or delaying the onset of the disease, and/or also lessening the severity or frequency of symptoms of the disease). The amount that will be therapeutically effective in the treatment of a particular individual's disorder or condition will depend on the symptoms and severity of the disease, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the disease or disorder, and should be decided according to the judgment of a practitioner and each patient's circumstances. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

In one embodiment, a nucleic acid of the invention (*e.g.*, a nucleic acid encoding an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, such as SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9, which may optionally comprise at least one polymorphism, or a nucleic acid that encodes an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide or a variant, derivative or fragment thereof, such as a nucleic acid encoding the protein of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10) can be used, either alone or in a pharmaceutical composition as described above. For example, *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* or a cDNA encoding an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, either by itself or included within a vector, can be introduced into cells (either *in vitro* or *in vivo*) such that the cells produce native HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide. If desired, cells that have been transformed with the gene or cDNA or a vector comprising the gene or cDNA can be introduced (or re-introduced) into an individual affected with the disease. Thus, cells that, in nature, lack native *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* expression and activity, or have mutant *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* expression and activity, or have expression of a disease-associated HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) variant,

can be engineered to express an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide or an active fragment of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide (or a different variant of an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide). In a preferred embodiment, nucleic acid encoding the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide, or an active fragment or derivative thereof, can be introduced into an expression vector, such as a viral vector, and the vector can be introduced into appropriate cells in an animal. Other gene transfer systems, including viral and nonviral transfer systems, can be used. Alternatively, nonviral gene transfer methods, such as calcium phosphate coprecipitation, mechanical techniques (*e.g.*, microinjection); membrane fusion-mediated transfer via liposomes; or direct DNA uptake, can also be used to introduce the desired nucleic acid molecule into a cell.

Alternatively, in another embodiment of the invention, a nucleic acid of the invention; a nucleic acid complementary to a nucleic acid of the invention; or a portion of such a nucleic acid (*e.g.*, an oligonucleotide as described below), can be used in "antisense" therapy, in which a nucleic acid (*e.g.*, an oligonucleotide) that specifically hybridizes to the RNA and/or genomic DNA of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* is administered or generated *in situ*. The antisense nucleic acid that specifically hybridizes to the RNA and/or DNA inhibits expression of the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule, *e.g.*, by inhibiting translation and/or transcription. Binding of the antisense nucleic acid can be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interaction in the major groove of the double helix.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid as described above. When the plasmid is transcribed in the cell, it produces RNA that is complementary to a portion of the mRNA and/or DNA that encodes an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide. Alternatively, the antisense construct can be an oligonucleotide probe which is generated *ex vivo* and introduced

into cells; it then inhibits expression by hybridizing with the mRNA and/or genomic DNA of *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. In one embodiment, the oligonucleotide probes are modified oligonucleotides that are resistant to endogenous nucleases, *e.g.* exonucleases and/or endonucleases, thereby rendering them stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Patent Nos. 5,176,996; 5,264,564; and 5,256,775). Additionally, general approaches to constructing oligomers useful in antisense therapy are also described, for example, by Van der Krol *et al.*, Biotechniques 6: 958-976 (1988); and Stein *et al.*, Cancer Res 48: 2659-2668 (1988). With respect to antisense DNA, oligodeoxyribonucleotides derived from the translation initiation site, *e.g.* between the -10 and +10 regions of an *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid sequence, are preferred.

To perform antisense therapy, oligonucleotides (RNA, cDNA or DNA) are designed that are complementary to mRNA encoding an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide. The antisense oligonucleotides bind to *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* mRNA transcripts and prevent translation. Absolute complementarity, although preferred, is not required. A sequence "complementary" to a portion of an RNA, as referred to herein, indicates that a sequence has sufficient complementarity to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The ability to hybridize will depend on both the degree of complementarity and the length of the antisense nucleic acid, as described in detail above. Generally, the longer the hybridizing nucleic acid, the more base mismatches with an RNA it may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures.

The oligonucleotides used in antisense therapy can be DNA, RNA, or chimeric mixtures or derivatives or modified versions thereof, single-stranded or double-stranded. The oligonucleotides can be modified at the base moiety, sugar

moiety, or phosphate backbone, for example, to improve stability of the molecule, hybridization, etc. The oligonucleotides can include other appended groups such as peptides (*e.g.* for targeting host cell receptors *in vivo*), or compounds facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, Proc. Natl. Acad. Sci. USA 86: 6553-6556 (1989); Lemaitre *et al.*, Proc. Natl. Acad Sci. USA 84: 648-652 (1987); PCT International Publication No. W088/09810)) or the blood-brain barrier (see, *e.g.*, PCT International Publication No. W089/10134), or hybridization-triggered cleavage agents (see, *e.g.*, Krol *et al.*, BioTechniques 6: 958-976 (1988)) or intercalating agents. (See, *e.g.*, Zon, Pharm. Res. 5: 539-549 (1988)). To this end, the oligonucleotide may be conjugated to another molecule (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent).

The antisense molecules are delivered to cells that express *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS) in vivo*. A number of methods can be used for delivering antisense DNA or RNA to cells; *e.g.*, antisense molecules can be injected directly into the tissue site, or modified antisense molecules, designed to target the desired cells (*e.g.*, antisense linked to peptides or antibodies that specifically bind receptors or antigens expressed on the target cell surface) can be administered systematically. Alternatively, in a preferred embodiment, a recombinant DNA construct is utilized in which the antisense oligonucleotide is placed under the control of a strong promoter (*e.g.*, pol III or pol II). The use of such a construct to transfect target cells in the patient results in the transcription of sufficient amounts of single stranded RNAs that will form complementary base pairs with the endogenous *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* transcripts and thereby prevent translation of the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* mRNA. For example, a vector can be introduced *in vivo* such that it is taken up by a cell and directs the transcription of an antisense RNA. Such a vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art and described above. For example, a plasmid, cosmid, YAC, or viral vector can be used to prepare the recombinant DNA

construct that can be introduced directly into the tissue site. Alternatively, viral vectors can be used that selectively infect the desired tissue, in which case administration may be accomplished by another route (*e.g.*, systematically).

Endogenous *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS),* or *HDRP(ΔNLS)* expression can also be reduced by inactivating or "knocking out" *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid sequences or their promoters using targeted homologous recombination (*e.g.*, see Smithies *et al.*, Nature 317: 230-234 (1985); Thomas and Capecchi, Cell 51: 503-512 (1987); Thompson *et al.*, Cell 5: 313-321 (1989)). For example, a mutant, non-functional *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* (either the coding regions or regulatory regions of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS) in vivo*. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. The recombinant DNA constructs can be directly administered or targeted to the required site *in vivo* using appropriate vectors, as described above. Alternatively, expression of non-mutant *HDAC9*, *HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* can be increased using a similar method: Targeted homologous recombination can be used to insert a DNA construct comprising a non-mutant, functional *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* (*e.g.*, a gene having SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9, which may optionally comprise at least one polymorphism), or a portion thereof, in place of a mutant *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* in the cell, as described above. In another embodiment, targeted homologous recombination can be used to insert a DNA construct comprising a nucleic acid that encodes an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide variant that differs from that present in the cell.

Alternatively, endogenous *HDAC9*, *HDAC9a*, *HDAC9*(ΔNLS), *HDAC9a*(ΔNLS), or *HDRP*(ΔNLS) expression can be reduced by targeting deoxyribonucleotide sequences complementary to the regulatory region of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* (*i.e.*, the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* promoter and/or enhancers) to form triple helical structures that prevent transcription of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* in target cells in the body. (See generally, Helene Anticancer Drug Des., 6(6): 569-84 (1991); Helene *et al.*, Ann, N.Y. Acad. Sci., 660: 27-36 (1992); and Maher, Bioassays 14(12): 807-15 (1992)). Likewise, the antisense constructs described herein, by antagonizing the normal biological activity of one of the HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) proteins, can be used in the manipulation of tissue, *e.g.*, tissue differentiation, both *in vivo* and for *ex vivo* tissue cultures. Furthermore, the antisense techniques (*e.g.*, microinjection of antisense molecules, or transfection with plasmids whose transcripts are anti-sense with regard to an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* mRNA or gene sequence) can be used to investigate role of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) in developmental events, as well as the normal cellular function of HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) in adult tissue. Such techniques can be utilized in cell culture, but can also be used in the creation of transgenic animals.

In yet another embodiment of the invention, other HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) therapeutic compounds as described herein can also be used in the treatment or prevention of a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. The therapeutic compounds can be delivered in a composition, as described above, or by themselves. They can be administered systemically, or can be targeted to a particular tissue. The therapeutic compounds can be produced by a variety of means, including chemical synthesis; recombinant production; *in vivo* production (*e.g.*, a transgenic animal, such as U.S. Patent No. 4,873,316 to Meade *et al.*), for example, and can be isolated using standard means such as those described herein.

A combination of any of the above methods of treatment (*e.g.*, administration of non-mutant HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide in conjunction with antisense therapy targeting mutant *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* mRNA; administration of a first variant encoded by *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* in conjunction with antisense therapy targeting a second encoded by *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, can also be used.

In another embodiment, the invention is directed to *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecules and HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptides for use as a medicament in therapy. For example, the nucleic acid molecules or polypeptides of the present invention can be used in the treatment of a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. In addition, the *HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecules and HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptides described herein can be used in the manufacture of a medicament for the treatment of a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

The invention will be further described by the following non-limiting examples. The teachings of all publications cited herein are incorporated herein by reference in their entirety.

EXEMPLIFICATION

Cloning of cDNA encodes a novel HDAC, designated HDAC9

HDAC9 was cloned by PCR and 3' rapid amplification of cDNA ends using primers designed from the sequence of human chromosome 7 whose translated product exhibited 80% identity to the HDAC domain of HDAC4, described in detail as follows.

Database analyses indicate that *HDRP* is located on chromosome 7 (7p15-p21). The human genome database (February 2001 release) of GenBank was searched using the human HDAC4 amino acid sequence. The TBLASTN program

was used to identify open reading frames downstream of *HDRP* on chromosome 7 that exhibit significant homology to the HDAC domain of HDAC4. Several fragments whose translated products exhibit over 58% identity were retrieved. Two sense primers (OL486, 5'-CCATGGAAACGGTACCCAGCAGGC-3' (SEQ ID NO: 16) and OL487, 5'-CACTCCATCGCTATGATGAAGGG-3' (SEQ ID NO: 17)) and antisense primers (OL484, 5'-AGTTCCCTTCATCATAGCGATGG-3' (SEQ ID NO: 18) and OL485, 5'-AATGTACAGGATGCTGGGGT-3' (SEQ ID NO: 19)) each were designed based upon one of these fragments whose translated products matched amino acids 842-873 of HDAC4. RT-PCR was performed using each of the antisense primers and a sense primer (5'-CCCTTGTAGCTGGTGGAGTTCCCTT-3' (SEQ ID NO: 20)) from the coding region of *HDRP* and human brain cDNA as a template. PCR was performed in a Biometra TGRADIENT Thermocycler for 30 cycles at 95°C for 20 seconds, 60°C for 20 seconds, and 72°C for 120 seconds.

3'-rapid amplification of cDNA ends was performed using the sense primer OL486 and adaptor primer 1 (Clontech), and marathon-ready cDNA from human brain (Clontech, Palo Alto, CA) according to the manufacturer's instruction. The products were re-amplified using nested sense primer OL487 and adaptor primer 2 (Clontech, Palo Alto, CA). PCR products were cloned into pGEM-T-easy vector (Promega, Madison, WI) and sequenced using an automated DNA sequencer at the DNA Sequencing Core Facility of the Memorial Sloan-Kettering Cancer Center, using DNA sequencing methods known to one of skill in the art.

Two cDNAs were cloned from the above-described methods. One cDNA (SEQ ID NO:1) encodes an HDAC9 protein that is 1011 amino acids in length. The other cDNA (SEQ ID NO: 3) encodes an HDAC9a protein that is 879 amino acids long. The cDNA sequence and amino sequence of *HDAC9* and *HDAC9a* are shown in FIGS. 1A-1G and FIGS. 2A-2B, respectively. Database analyses of these cDNAs against human genomic DNA sequences indicated that these two cDNAs are generated by alternatively splicing. An alignment of HDAC9, HDAC9a, HDRP, and HDAC4 is shown in FIGS. 3A-3C.

Each of the HDAC9 and HDAC9a nucleic acid sequences were cloned into the pFLAG-CMV-5b vector (Sigma) in frame with the C-terminal FLAG tag. Only

the coding regions plus three extra base pairs (ACC) of cDNA of the HDAC9 and HDAC9a nucleic acid sequences were included in the constructs. These constructs are referred to herein as HDAC9-FLAG and HDAC9a-FLAG, respectively. These constructs are contained in *E. coli*, and can readily be expressed. For HDAC9, the insert is 3033 bp and for HDAC9a, the insert size is 2637 bp. Both HDAC9 and HDAC9a can be released with EcoRV and BamHI (whose sites have been incorporated in the primers to obtain HDAC9 and HDAC9a coding cDNA for cloning purpose) restriction enzyme digestion.

The *HDAC9* cDNA sequences from the known 5'-end of *HDRP* cDNA to the 3'-untranslated region cloned in this study cover over 511 kb of genomic DNA on chromosome 7. As shown in FIG. 4, the coding region cDNA of *HDAC9* resides in 23 exons spanning 458 kb of genomic sequence. Exons 21, 22, and 23 are one single exon in HDAC9a, but the middle exon that is numbered exon 22 in FIG. 4, containing an in-frame stop codon, is spliced out in HDAC9. In addition, exons 12 and 13 are a single exon used by HDRP. Exon 13 is spliced as part of an intron in HDAC9 and HDAC9a.

Further analysis revealed that exon 7, which contains a nuclear localization signal (NLS) is alternatively spliced in an HDRP isoform, creating HDRP(ΔNLS). RT-PCR analyses using primers based on sequences from exon 6 and exon 14 indicate that this alternative splicing event also occurs in *HDAC9* and/or *HDAC9a*. Thus, it is possible that at least 6 proteins can be generated from a single *HDAC9* gene by alternatively splicing of its RNA. The cDNA sequences and amino acid sequences for HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), and HDRP(ΔNLS) are shown in FIGS. 1A-1O and 2A-2E, respectively.

HDAC9 mRNA is differentially expressed among human tissues

The expression of *HDAC9* mRNA was determined by Northern blot analysis using a human multiple tissue Northern blot (Clontech, Palo Alto, CA). Hybridization was performed according to the manufacturer's instruction using ExPressHyb solution (Clontech, Palo Alto, CA). The ^{32}P-random priming labeled 3'-untranslated region common to both *HDAC9* and *HDAC9a* that shares no significant sequence homology with *HDRP* was used as a probe. Two transcripts at

9.8 and 4.1 kb were detected in all tissues examined (FIG. 6A). The 4.1 kb transcript is shorter than the 4.4 kb *HDRP* transcript (See Zhou, *et al.*, Proc. Natl. Acad. Sci. USA, 97:1056-1061 (2000)). A third transcript at 1.2 kb was detected in placenta (FIG. 6A). Similar to *HDRP* (See Zhou, X., *et al.*, Proc. Natl. Acad. Sci. USA, 97:1056-1061 (2000)), high levels of *HDAC9* transcripts were detected in brain and skeletal muscle (FIG. 6A).

The distribution of alternatively spliced mRNA variants among tissues was examined by RT-PCR using primers (OL516 5'-TGTGTCATCGAGCTGGCTTC-3' (SEQ ID NO: 21) and OL517 5'-ATCTTCTGCAAGTGGCTCCA-3' (SEQ ID NO: 22)) spanning the alternatively spliced exon 22 and cDNA panel from the same tissues as the multiple tissue Northern blot. PCR was performed in a Biometra TGRADIENT Thermocycler for 30 cycles at 95°C for 20 seconds, 60°C for 20 seconds, and 72°C for 60 seconds. The expected sizes of PCR products were 680 base pairs for *HDAC9* and 993 base pairs for *HDAC9a*. The ratio of *HDAC9* and *HDAC9a* transcripts differed among tissues (FIG. 6B). In the placenta and kidney, the levels of the two transcripts were about the same (FIG. 6B). In the brain, heart, and pancreas, there were more transcripts of *HDAC9* than *HDAC9a*. In the other tissues examined, there were more *HDAC9a* transcripts than *HDAC9* transcripts (FIG. 6B). Under the conditions tested, *HDAC9* transcripts were undetectable in liver (FIG. 6B). The lung had an *HDAC9* product that was larger than expected and abundant. The lung also had low levels of *HDAC9* transcripts and *HDAC9a* transcripts (FIG. 6B). An additional PCR product was also amplified from cDNA of the pancreas; this product was than the expected products from *HDAC9* and *HDAC9a* (FIG. 6B). The identity of the different sized transcripts is unknown.

HDAC9 and HDAC9a possess histone deacetylase activity

HDAC9 was named based on sequence homology to HDAC4 (FIGS. 3A-3C). To determine whether HDAC9 and HDAC9a possess HDAC activity, an HDAC enzymatic assay was performed using anti-FLAG immunoprecipitated HDAC9-FLAG and HDAC9a-FLAG.

C-terminal FLAG-tagged HDAC9 (HDAC9-FLAG) and HDAC9a (HDAC9a-FLAG) expression vectors were constructed using the pFLAG-CMV-5b

vector (Sigma) and PCR amplified coding regions of HDAC9 and HDAC9a in frame with the FLAG-tag to form pFLAG-CMV-5b-HDAC9 (plasmid VR1) and pFLAG-CMV-5b-HDAC9a (plasmid VR2). All constructs were confirmed by DNA sequencing.

Transfection of human kidney 293T cells, immunoprecipitation using anti-FLAG M2 Agarose (Sigma), Western blot analyses and dual luciferase assays were performed essentially as previously described by Zhou *et al.* (Proc. Natl. Acad. Sci. USA, 97:1056-1061 (2000)). Briefly, the cells (American Type Culture Collection) were cultured in DME HG medium (GIBCO/BRL) supplemented with 10% (vol/vol) FBS at 37 °C in a 5% CO_2 atmosphere. Transient transfection was performed by using Lipofectamine (GIBCO/BRL) or Fugene 6 (Roche Molecular Biochemicals) according to the manufacturers' instructions. Cells were harvested 24 to 48 hours after transfection and lysed in IP lysis buffer (50 mM Tris·HCl, pH 7.5/120 mM NaCl/5 mM EDTA/0.5% NP-40) at 5 x 10^7 cells per ml. Immunoprecipitation with anti-FLAG M2-agarose (Sigma, St. Louis, MO) was performed according to the manufacturer's instructions. Immunoprecipitated proteins were released from the agarose beads by using FLAG-peptide and either used directly for HDAC enzymatic activity assays or resolved on SDS/PAGE for Western blot analyses. Anti-FLAG antibody was purchased from Sigma (St. Louis, MO). Western blot analyses were performed using standard methods.

HDAC9 and HDAC9a enzymatic activity were assessed with the HDAC Fluorescent Activity Assay/Drug Discovery Kit-AK-500 (BIOMOL Research Laboratories) using a FLUOR DE LYS™ that contains an acetylated lysine side chain as a substrate and immunoprecipitated HDAC9-FLAG and HDAC9a-FLAG polypeptides according to the manufacturer's instruction and a SPECTRAmax® GEMINI XS microplate spectrofluorometer using the SOFTmax® PRO system (Molecular Devices) at excitation 355 nm and emission 460 nm with a cut off filter of 455 nm. Briefly, HDAC9-FLAG and HDAC9a-FLAG were incubated with the substrate overnight at room temperature in a 96-well plate. The reaction was stopped by addition of Fluor De Lys™ Developer and samples were read with the fluorometer.

As shown in FIG. 7, both HDAC9-FLAG and HDAC9a-FLAG deacetylated the acetylated lysine of FLUOR DE LYS™ and the activity of HDAC9 and HDAC9a was comparable. To examine the activity of HDAC9 and HDAC9a, inhibition studies using TSA were carried out by preincubating HDAC9-FLAG and HDAC9a-FLAG with TSA for 15 minutes at room temperature. The assay was then carried out as stated above. As shown in FIG. 7, TSA inhibited HDAC9 and HDAC9a deacetylase activity. The inset gel in FIG. 7 shows the amount of protein used in the assay. SAHA, a potent HDAC inhibitor (Richon *et al.*, Proc. Natl. Acad. Sci. USA, 95:3003-3007 (1998)) also completely inhibited the histone deacetylase activity of HDAC9-FLAG and HDAC9a-FLAG. The HDAC activity of HDAC9 and HDAC9a was about ten times lower than the deacetylase activity of HDAC4 when comparable amount of protein was used under conditions tested here.

HDAC9 and HDAC9a enzymatic activity was also determined through HDAC enzymatic assays using ^{3}H-histones isolated from murine erythroleukemia cells as a substrate. This assay was performed essentially as described by Richon *et al.* (Proc. Natl. Acad. Sci. USA, 95:3003-3007 (1998)). Briefly, HDAC9-FLAG and HDAC9a-FLAG were incubated with ^{3}H-histones overnight at 37°C. The reaction was stopped by the addition of 1M HCl/0.1 acetic acid. Released ^{3}H-acetic acid was extracted with ethyl acetate and quantified by scintillation counting. For inhibition studies, the immunoprecipitated complexes were preincubated with the different HDAC inhibitors for 30 minutes at 4°C.

As shown in FIG. 8, HDAC9a-FLAG deacetylated ^{3}H-acetyl-histones. SAHA, a potent HDAC inhibitor also completely inhibited the histone deacetylase activity of HDAC9a-FLAG. TSA also inhibited HDAC9a deacetylase activity. Similar results were obtained when HDAC9 was used as the enzyme source.

HDAC9 and HDAC9a repress MEF2-mediated transcription

The Xenopus homolog of HDRP, MITR, was identified as a MEF2 interacting transcriptional repressor (Sparrow *et al.*, EMBO J. 18:5085-5098(1999)) and mouse HDRP also interacts with and represses MEF2 mediated transcription (Zhang *et al.*, J. Biol. Chem. 276:35-39 (2001)). We first tested whether HDAC9-FLAG and HDAC9a-FLAG interact with MEF2. 293 cells were transfected with

vector, HDAC9-FLAG, or HDAC9a-FLAG. The cells were subsequently lysed and HDAC9-FLAG and HDAC9a-FLAG proteins were immunoprecipitated with anti-FLAG antibodies. Western blot analysis of the immunoprecipitated proteins was carried out, using anti-MEF-2 antibody to probe the blot. As shown in FIG. 9A, both HDAC9 and HDAC9a interacted with MEF2 in 293T cells.

It was then determined whether HDAC9 and HDAC9a repress MEF2-mediated transcription. This determination was carried out as follows. The p3XMEF2-luciferase reporter gene (100 ng) and the vector pRL-TK (Promega) (5 ng) were co-transfected into 293T cells in the absence (pcDNA3 empty vector) or presence of MEF2C (100 ng of pCMV-MEF2C). HDAC9-F (1 ng, 10 ng, or 100 ng of pFLAG-HDAC9; pFLAG-HDAC9 and HDAC9-FLAG are different constructs, with the FLAG sequence located at opposite ends of the HDAC9 nucleotide, but are functionally equivalent) or HDAC9a-F (1 ng, 10 ng, or 100 ng of pFLAG-HDAC9a; pFLAG-HDAC9a and HDAC9a-FLAG are different constructs, with the FLAG sequence located at opposite ends of the HDAC9a nucleotide, but are functionally equivalent) was included in a subset of experimental groups with the MEF2C vector. pFLAG empty vector was used to adjust the DNA to an equal amount in each transfection. The cells were harvested 24 to 36 hours after transfection and the luciferase activities were measured using the Dual-Luciferase™ Reporter Assay System from Promega according to the manufacturer's instruction. The firefly luciferase activity was first normalized to the co-transfected Renilla luciferase activity (encoded by the pRL-TK vector), and the luciferase activity value for cells transfected with MEF2C alone was set at 1. MEF2C activated transcription over 30 times the basal level of transcription. As shown in FIG. 9B, HDAC9-FLAG and HDAC9a-FLAG repressed MEF2C mediated transcriptional activation in a dose-dependent manner and completely abolished the activation at the 100 ng dose for both HDAC9 and HDAC9a. The transcriptional repression effect of HDAC9 and HDAC9a on MEF2C mediated transcription was a specific effect since a co-transfected reporter gene for transfection efficiency containing a TK promoter was not repressed by HDAC9 or HDAC9a.

Described herein is the identification and characterization of a new class II HDAC, designated HDAC9. HDAC9 has several alternatively spliced isoforms,

one of which is the previously identified HDRP (Zhou *et al.*, Proc. Natl. Acad. Sci. USA 97:1056-1061 (2000)). HDAC9 and HDAC9a possess HDAC activity, which appears to have a lower specific enzymatic activity than HDAC4. While not wishing to be bound by any particular theory, it is possible that an essential co-factor is lost during immunoprecipitation or does not exist in 293T cells (for example, metastasis-associated protein 2 is essential for the assembly of a catalytically active HDAC1 (Zhang *et al.*, Genes Dev. 13:1924-1935 (1999)), the substrates used are not its natural substrate, or the FLAG tag which interferes with the folding of the protein.

Searching the human genome with the HDAC domain from either HDAC1 or HDAC9 identified a total of 10 HDACs in the presently completed human genome sequence, a number of which are schematically represented in FIG. 10. HDACs 1, 2, 3, 8, 4, 5, 6, 7, 9, and 9a all have HDAC domains. HDRP, which is also schematically depicted in FIG. 10, does not have a catalytic domain.

All references described herein are incorporated by reference in their entirety. While this invention has been particularly shown and described with reference to preferred embodiment thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the appended claims.

CLAIMS

What is claimed is:

1. An isolated or recombinant histone deacetylase polypeptide, said polypeptide selected from:
   a) an isolated or recombinant polypeptide comprising SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; and
   b) an isolated or recombinant polypeptide having at least 60% sequence identity with any one of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

2. The isolated or recombinant histone deacetylase polypeptide of Claim 1, said polypeptide selected from:
   a) a polypeptide consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

3. The isolated or recombinant histone deacetylase polypeptide of Claim 1, wherein said polypeptide is human.

4. An isolated nucleic acid molecule selected from the group:
   a) an isolated nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9;
   b) a complement of an isolated nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9
   c) an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;

d) a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;

e) a nucleic acid that is hybridizeable under high stringency conditions to a nucleic acid molecule that encodes any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, or SEQ ID NO: 8, or a complement thereof; or

f) a nucleic acid molecule that is hybridizeable under high stringency conditions to a nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, or SEQ ID NO: 7; and

g) an isolated nucleic acid molecule that has at least 55% sequence identity with any one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, or a complement thereof.

5. The isolated nucleic acid molecule of Claim 4, said nucleic acid molecule consisting of the nucleic acid molecule selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 9.

6. The isolated nucleic acid molecule of Claim 4, wherein said nucleic acid molecule is human.

7. A vector comprising the isolated nucleic acid molecule of Claim 4.

8. A cell comprising the vector of Claim 7.

9. A cell comprising the isolated nucleic acid molecule of Claim 4.

10. A purified antibody that selectively binds a polypeptide of Claim 1.

11. A method of identifying a compound that modulates expression of a nucleic acid molecule of Claim 4, said method comprising the steps of:

a) contacting said nucleic acid molecule with a candidate compound under conditions suitable for expression; and

b) assessing the level of expression of said nucleic acid molecule,

wherein a candidate compound that increases or decreases expression of said nucleic acid molecule relative to a control is a compound that modulates expression of said nucleic acid molecule.

12. The method of Claim 11, wherein said method is carried out in a cell or animal.

13. The method of Claim 11, wherein said method is carried out in a cell free system.

14. A method of identifying a compound that modulates the enzymatic activity of the polypeptide of Claim 1, said method comprising the steps of:

a) contacting said polypeptide with a candidate compound under conditions suitable for enzymatic reaction; and

b) assessing the enzymatic activity level of said polypeptide,

wherein a candidate compound that increases or decreases the enzymatic activity level of said polypeptide relative to a control is a compound that modulates the enzymatic activity of said polypeptide.

15. The method of Claim 14, wherein said method is carried out in a cell or animal.

16. The method of Claim 14, wherein said method is carried out in a cell free system.

17. The method of Claim 14, wherein said polypeptide is further contacted with a substrate for the polypeptide, and wherein said substrate is selected from the group consisting of a cell proliferation disease binding agent, an

apoptotic disease binding agent, and a cell differentiation disease binding agent.

18. The method of Claim 17, wherein said candidate compound is an inhibitor.

19. The method of Claim 17, wherein said candidate compound is an activator.

20. A method of identifying a compound that modulates the transcriptional repression activity of the polypeptide of Claim 1, said method comprising the steps of:
    a) contacting said polypeptide with a candidate compound under conditions suitable for a transcriptional repression reaction; and
    b) assessing the transcriptional repression activity level of said polypeptide,

wherein a candidate compound that increases or decreases the transcriptional repression activity level of said polypeptide relative to a control is a compound that modulates the transcriptional repression activity of said polypeptide.

21. The method of Claim 20, wherein said method is carried out in a cell or animal.

22. The method of Claim 20, wherein said method is carried out in a cell free system.

23. The method of Claim 20, wherein said polypeptide is further contacted with a substrate for the polypeptide, and wherein said substrate is selected from the group consisting of a cell proliferation disease binding agent, an apoptotic disease binding agent, and a cell differentiation disease binding agent.

24. The method of Claim 23, wherein said candidate compound is an inhibitor.

25. The method of Claim 23, wherein said candidate compound is an activator.

26. A method of identifying a compound that modulates expression of a nucleic acid molecule of Claim 4, said method comprising the steps of:

a) providing a nucleic acid molecule comprising a promoter region of said nucleic acid of Claim 4 or part of a promoter region of said nucleic acid of Claim 4 operably linked to a reporter gene;

b) contacting said nucleic acid molecule or with a candidate compound; and

c) assessing the level of said reporter gene,

wherein a candidate compound that increases or decreases expression of said reporter gene relative to a control is a compound that modulates expression of said nucleic acid molecule of Claim 4.

27. The method of Claim 26, wherein said method is carried out in a cell.

28. A method of identifying a polypeptide that interacts with a polypeptide of Claim 1 in a yeast two-hybrid system, said method comprising the steps of:

a) providing a first nucleic acid vector comprising a nucleic acid molecule encoding a DNA binding domain and said polypeptide of Claim 1;

b) providing a second nucleic acid vector comprising a nucleic acid encoding a transcription activation domain and a nucleic acid encoding a test polypeptide;

c) contacting said first nucleic acid vector with said second nucleic acid vector in a yeast two-hybrid system; and

d) assessing transcriptional activation in said yeast two-hybrid system,

wherein an increase in transcriptional activation relative to a control indicates that the test polypeptide is a polypeptide that interacts with said polypeptide of Claim 1.

29. A pharmaceutical composition comprising a polypeptide of Claim 1.

30. A method of diagnosing a cell proliferation disease, an apoptotic disease, or a cell differentiation disease in a subject, said method comprising the steps of:

a) obtaining a sample from said subject; and

b) assessing the level of activity or expression of said polypeptide of Claim 1 in said sample, or detecting the level of said nucleic acid molecule of Claim 4,

wherein if said level is increased relative to a control, then said subject has an increased likelihood of having a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, and wherein if said level is decreased relative to a control, then said subject has a decreased likelihood of having a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

31. The method of Claim 30, wherein said level of activity or expression of said polypeptide of Claim 1 in said sample is measured using immunohistochemical techniques.

32. The method of Claim 30, wherein said level of said nucleic acid molecule of Claim 4 in said sample is measured using *in situ* hybridization techniques.

33. A method of treating a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, said method comprising administering a compound identified by the method of Claim 14.

34. A method of treating a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, said method comprising administering a compound identified by the method of Claim 20.

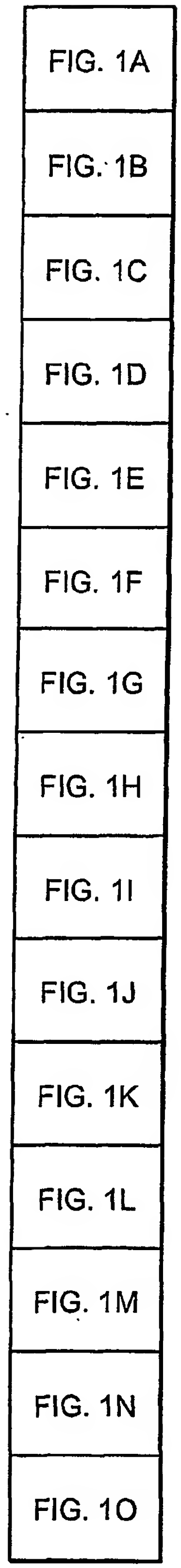


FIG. 1A
FIG. 1B
FIG. 1C
FIG. 1D
FIG. 1E
FIG. 1F
FIG. 1G
FIG. 1H
FIG. 1I
FIG. 1J
FIG. 1K
FIG. 1L
FIG. 1M
FIG. 1N
FIG. 1O

FIG. 1

HDAC93186 bp Coding 151-3186

```
      Exon 1
1     ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact

                                                   2
101   ctaagccaga tggggtggct ggacgagagc agctcttggc tcagcaaaga ATGCACAGTA TGATCAGCTC AGTGGATGTG AAGTCAGAAG TTCCTGTGGG

201   CCTGGAGCCC ATCTCACCTT TAGACCTAAG GACAGACCTC AGGATGATGA TGCCCGTGGT GGACCCTGTT GTCCGTGAGA AGCAATTGCA GCAGGAATTA

                            3
301   CTTCTTATCC AGCAGCAGCA ACAAATCCAG AAGCAGCTTC TGATAGCAGA GTTTCAGAAA CAGCATGAGA ACTTGACACG GCAGCACCAG GCTCAGCTTC

                                                              4
401   AGGAGCATAT CAAGGAACTT CTAGCCATAA AACAGCAACA AGAACTCCTA GAAAAGGAGC AGAAACTGGA GCAGCAGAGG CAAGAACAGG AAGTAGAGAG

501   GCATCGCAGA GAACAGCAGC TTCCTCCTCT CAGAGGCAAA GATAGAGGAC GAGAAAGGGC AGTGGCAAGT ACAGAAGTAA AGCAGAAGCT TCAAGAGTTC

                                                                         5
601   CTACTGAGTA AATCAGCAAC GAAAGACACT CCAACTAATG GAAAAAATCA TTCCGTGAGC CGCCATCCCA AGCTCTGGTA CACGGCTGCC CACCACACAT

                                         6
701   CATTGGATCA AAGCTCTCCA CCCCTTAGTG GAACATCTCC ATCCTACAAG TACACATTAC CAGGAGCACA AGATGCAAAG GATGATTTCC CCCTTCGAAA

                                                    7
801   AACTGCCTCT GAGCCCAACT TGAAGGTGCG GTCCAGGTTA AAACAGAAAG TGGCAGAGAG GAGAAGCAGC CCCTTACTCA GGCGGAAGGA TGGAAATGTT
```

FIG. 1A

```
                                                                  8
901  GTCACTTCAT TCAAGAAGCG AATGTTTGAG GTGACAGAAT CCTCAGTCAG TAGCAGTTCT CCAGGCTCTG GTCCCAGTTC ACCAAACAAT GGGCCAACTG
                                                                                   9
1001 GAAGTGTTAC TGAAAATGAG ACTTCGGTTT TGCCCCCTAC CCCTCATGCC GAGCAAATGG TTTCACAGCA ACGCATTCTA ATTCATGAAG ATTCCATGAA

1101 CCTGCTAAGT CTTTATACCT CTCCTTCTTT GCCCAACATT ACCTTGGGGC TTCCCGCAGT GCCATCCCAG CTCAATGCTT CGAATTCACT CAAAGAAAAG

1201 CAGAAGTGTG AGACGCAGAC GCTTAGGCAA GGTGTTCCTC TGCCTGGGCA GTATGGAGGC AGCATCCCGG CATCTTCCAG CCACCCTCAT GTTACTTTAG
                                                                        10
1301 AGGGAAAGCC ACCCAACAGC AGCCACCAGG CTCTCCTGCA GCATTTATTA TTGAAAGAAC AAATGCGACA GCAAAAGCTT CTTGTAGCTG GTGGAGTTCC

1401 CTTACATCCT CAGTCTCCCT TGGCAACAAA AGAGAGAATT TCACCTGGCA TTAGAGGTAC CCACAAATTG CCCCGTCACA GACCCCTGAA CCGAACCCAG
                                                           11
1501 TCTGCACCTT TGCCTCAGAG CACGTTGGCT CAGCTGGTCA TTCAACAGCA ACACCAGCAA TTCTTGGAGA AGCAGAAGCA ATACCAGCAG CAGATCCACA

1601 TGAACAAACT GCTTTCGAAA TCTATTGAAC AACTGAAGCA ACCAGGCAGT CACCTTGAGG AAGCAGAGGA AGAGCTTCAG GGGGACCAGG CGATGCAGGA
                                                      12
1701 AGACAGAGCG CCCTCTAGTG GCAACAGCAC TAGGAGCGAC AGCAGTGCTT GTGTGGATGA CACACTGGGA CAAGTTGGGG CTGTGAAGGT CAAGGAGGAA

1801 CCAGTGGACA GTGATGAAGA TGCTCAGATC CAGGAAATGG AATCTGGGGA GCAGGCTGCT TTTATGCAAC AGCCTTTCCT GGAACCCACG CACACACGTG
```

FIG. 1B

1901 CGCTCTCTGT GCGCCAAGCT CCGCTGGCTG CGGTTGGCAT GGATGGATTA GAGAAACACC GTCTCGTCTC CAGGACTCAC TCTTCCCCTG CTGCCTCTGT

14

2001 TTTACCTCAC CCAGCAATGG ACCGCCCCCT CCAGCCTGGC TCTGCAACTG GAATTGCCTA TGACCCCTTG ATGCTGAAAC ACCAGTGCGT TTGTGGCAAT

15

2101 TCCACCACCC ACCCTGAGCA TGCTGGACGA ATACAGAGTA TCTGGTCACG ACTGCAAGAA ACTGGGCTGC TAAATAAATG TGAGCGAATT CAAGGTCGAA

16

2201 AAGCCAGCCT GGAGGAAATA CAGCTTGTTC ATTCTGAACA TCACTCACTG TTGTATGGCA CCAACCCCCT GGACGGACAG AAGCTGGACC CCAGGATACT

17

2301 CCTAGGTGAT GACTCTCAAA AGTTTTTTTC CTCATTACCT TGTGGTGGAC TTGGGGTGGA CAGTGACACC ATTTGGAATG AGCTACACTC GTCCGGTGCT

18 19

2401 GCACGCATGG CTGTTGGCTG TGTCATCGAG CTGGCTTCCA AAGTGGCCTC AGGAGAGCTG AAGAATGGGT TTGCTGTTGT GAGGCCCCCT GGCCATCACG

20

2501 CTGAAGAATC CACAGCCATG GGGTTCTGCT TTTTTAATTC AGTTGCAATT ACCGCCAAAT ACTTGAGAGA CCAACTAAAT ATAAGCAAGA TATTGATTGT

21

2601 AGATCTGGAT GTTCACCATG GAAACGGTAC CCAGCAGGCC TTTTATGCTG ACCCCAGCAT CCTGTACATT TCACTCCATC GCTATGATGA AGGGAACTTT

23

2701 TTCCCTGGCA GTGGAGCCCC AAATGAGGTT GGAACAGGCC TTGGAGAAGG GTACAATATA AATATTGCCT GGACAGGTGG CCTTGATCCT CCCATGGGAG

24

2801 ATGTTGAGTA CCTTGAAGCA TTCAGGAcca TCGTGAAGCC TGTGGCCAAA GAGTTTGATC CAGACATGGT CTTAGTATCT GCTGGATTTG ATGCATTGGA

25

2901 AGGCCACACC CCTCCTCTAG GAGGGTACAA AGTGACGGCA AAATGTTTTG GTCATTTGAC GAAGCAATTG ATGACATTGG CTGATGGACG TGTGGTGTTG

3001 GCTCTAGAAG GAGGACATGA TCTCACAGCC ATCTGTGATG CATCAGAAGC CTGTGTAAAT GCCCTTCTAG GAAATGAGCT GGAGCCACTT GCAGAAGATA

26

3101 TTCTCCACCA AAGCCCGAAT ATGAATGCTG TTATTTCTTT ACAGAAGATC ATTGAAATTC AAAGTATGTC TTTAAAGTTC TCTTAA

FIG. 1C

HDAC9a 3499 bp (Coding 151-2790)

Exon 1

```
  1 ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact
                                 2
101 ctaagccaga tggggtggct ggacgagagc agctcttggc tcagcaaaga ATGCACAGTA TGATCAGCTC AGTGGATGTG AAGTCAGAAG TTCCTGTGGG

201 CCTGGAGCCC ATCTCACCTT TAGACCTAAG GACAGACCTC AGGATGATGA TGCCCGTGGT GGACCCTGTT GTCCGTGAGA AGCAATTGCA GCAGGAATTA
                                 3
301 CTTCTTATCC AGCAGCAGCA ACAAATCCAG AAGCAGCTTC TGATAGCAGA GTTTCAGAAA CAGCATGAGA ACTTGACACG GCAGCACCAG GCTCAGCTTC

401 AGGAGCATAT CAAGGAACTT CTAGCCATAA AACAGCAACA AGAACTCCTA GAAAAGGAGC AGAAACTGGA GCAGCAGAGG CAAGAACAGG AAGTAGAGAG
                                 4
501 GCATCGCAGA GAACAGCAGC TTCCTCCTCT CAGAGGCAAA GATAGAGGAC GAGAAAGGGC AGTGGCAAGT ACAGAAGTAA AGCAGAAGCT TCAAGAGTTC
```

FIG. 1D

```
                                            5
601  CTACTGAGTA AATCAGCAAC GAAAGACACT CCAACTAATG GAAAAAATCA TTCCGTGAGC CGCCATCCCA AGCTCTGGTA CACGGCTGCC CACCACACAT
                                            6
701  CATTGGATCA AAGCTCTCCA CCCCTTAGTG GAACATCTCC ATCCTACAAG TACACATTAC CAGGAGCACA AGATGCAAAG GATGATTTCC CCCTTCGAAA
                                            7
801  AACTGCCTCT GAGCCCAACT TGAAGGTGCG GTCCAGGTTA AAACAGAAAG TGGCAGAGAG GAGAAGCAGC CCCTTACTCA GGCGGAAGGA TGGAAATGTT

901  GTCACTTCAT TCAAGAAGCG AATGTTTGAG GTGACAGAAT CCTCAGTCAG TAGCAGTTCT CCAGGCTCTG GTCCCAGTTC ACCAAACAAT GGGCCAACTG
                                            8
1001 GAAGTGTTAC TGAAAATGAG ACTTCGGTTT TGCCCCCTAC CCCTCATGCC GAGCAAATGG TTTCACAGCA ACGCATTCTA ATTCATGAAG ATTCCATGAA
                                            9
1101 CCTGCTAAGT CTTTATACCT CTCCTTCTTT GCCCAACATT ACCTTGGGGC TTCCCGCAGT GCCATCCCAG CTCAATGCTT CGAATTCACT CAAAGAAAAG

1201 CAGAAGTGTG AGACGCAGAC GCTTAGGCAA GGTGTTCCTC TGCCTGGGCA GTATGGAGGC AGCATCCCGG CATCTTCCAG CCACCCTCAT GTTACTTTAG
                                            10
1301 AGGGAAAGCC ACCCAACAGC AGCCACCAGG CTCTCCTGCA GCATTTATTA TTGAAAGAAC AAATGCGACA GCAAAAGCTT CTTGTAGCTG GTGGAGTTCC

1401 CTTACATCCT CAGTCTCCCT TGGCAACAAA AGAGAGAATT TCACCTGGCA TTAGAGGTAC CCACAAATTG CCCCGTCACA GACCCCTGAA CCGAACCCAG
                                            11
1501 TCTGCACCTT TGCCTCAGAG CACGTTGGCT CAGCTGGTCA TTCAACAGCA ACACCAGCAA TTCTTGGAGA AGCAGAAGCA ATACCAGCAG CAGATCCACA
```

FIG. 1E

```
1601 TGAACAAACT GCTTTCGAAA TCTATTGAAC AACTGAAGCA ACCAGGCAGT CACCTTGAGG AAGCAGAGGA AGAGCTTCAG GGGGACCAGG CGATGCAGGA
                                  12
1701 AGACAGAGCG CCCTCTAGTG GCAACAGCAC TAGGAGCGAC AGCAGTGCTT GTGTGGATGA CACACTGGGA CAAGTTGGGG CTGTGAAGGT CAAGGAGGAA

1801 CCAGTGGACA GTGATGAAGA TGCTCAGATC CAGGAAATGG AATCTGGGGA GCAGGCTGCT TTTATGCAAC AGCCTTTCCT GGAACCCACG CACACACGTG
                                  14
1901 CGCTCTCTGT GCGCCAAGCT CCGCTGGCTG CGGTTGGCAT GGATGGATTA GAGAAACACC GTCTCGTCTC CAGGACTCAC TCTTCCCCTG CTGCCTCTGT

2001 TTTACCTCAC CCAGCAATGG ACCGCCCCCT CCAGCCTGGC TCTGCAACTG GAATTGCCTA TGACCCCTTG ATGCTGAAAC ACCAGTGCGT TTGTGGCAAT
                                  15
2101 TCCACCACCC ACCCTGAGCA TGCTGGACGA ATACAGAGTA TCTGGTCACG ACTGCAAGAA ACTGGGCTGC TAAATAAATG TGAGCGAATT CAAGGTCGAA
                                  16
2201 AAGCCAGCCT GGAGGAAATA CAGCTTGTTC ATTCTGAACA TCACTCACTG TTGTATGGCA CCAACCCCCT GGACGGACAG AAGCTGGACC CCAGGATACT
                                  17
2301 CCTAGGTGAT GACTCTCAAA AGTTTTTTTC CTCATTACCT TGTGGTGGAC TTGGGGTGGA CAGTGACACC ATTTGGAATG AGCTACACTC GTCCGGTGCT
                                  18                                                    19
2401 GCACGCATGG CTGTTGGCTG TGTCATCGAG CTGGCTTCCA AAGTGGCCTC AGGAGAGCTG AAGAATGGGT TTGCTGTTGT GAGGCCCCCT GGCCATCACG
                                  20
2501 CTGAAGAATC CACAGCCATG GGGTTCTGCT TTTTTAATTC AGTTGCAATT ACCGCCAAAT ACTTGAGAGA CCAACTAAAT ATAAGCAAGA TATTGATTGT
```

FIG. 1F

```
                                             21
2601 AGATCTGGAT GTTCACCATG GAAACGGTAC CCAGCAGGCC TTTTATGCTG ACCCCAGCAT CCTGTACATT TCACTCCATC GCTATGATGA AGGGAACTTT

2701 TTCCCTGGCA GTGGAGCCCC AAATGAGGTT CGGTTTATTT CTTTAGAGCC CCACTTTTAT TTGTATCTTT CAGGTAATTG CATTGCATGA  ttacccctaa
                                                                                               STOP CODON
                                             22
2801 ttttcttgtc ctttgctggt gttttaaatt acacgagatt actgaattgt cccatgggac caagaaccag tgcagaacaa gtgcataacc cagagcactg

2901 tttgtcaggg aaggttgggc tgatttgatg tgttgtttga tgtttatttc aagagctccc atgtgcttgt tttcctctct tcttgctttc ttccatttgc
                                             23
3001 tctcttctct gcccaccgtg gtgtgtcttt ctcttcccag gttggaacag gccttggaga agggtacaat ataaatattg cctggacagg tggccttgat

3101 cctcccatgg gagatgttga gtaccttgaa gcattcagga ccatcgtgaa gcctgtggcc aaagagtttg atccagacat ggtcttagta tctgctggat
                                             24
3201 ttgatgcatt ggaaggccac acccctcctc taggagggta caaagtgacg gcaaaatgtt ttggtcattt gacgaagcaa ttgatgacat tggctgatgg
                                             25
3301 acgtgtggtg ttggctctag aaggaggaca tgatctcaca gccatctgtg atgcatcaga agcctgtgta aatgcccttc taggaaatga gctggagcca
                                             26
3401 cttgcagaag atattctcca ccaaagcccg aatatgaatg ctgttatttc tttacagaag atcattgaaa ttcaaagtat gtctttaaag ttctcttaa
```

FIG. 1G

>HDRP (deltaNLS)

```
   1 ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca
  51 cttgcaggac tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact
 101 ctaagccaga tggggtggct ggacgagagc agctcttggc tcagcaaaga
 151 atgcacagta tgatcagctc agtggatgtg aagtcagaag ttcctgtggg
 201 cctggagccc atctcacctt tagacctaag gacagacctc aggatgatga
 251 tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta
 301 cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga
 351 gtttcagaaa cagcatgaga acttgacacg gcagcaccag gctcagcttc
 401 aggagcatat caaggaactt ctagccataa aacagcaaca agaactccta
 451 gaaaaggagc agaaactgga gcagcagagg caagaacagg aagtagagag
 501 gcatcgcaga gaacagcagc ttcctcctct cagaggcaaa gatagaggac
 551 gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc
 601 ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca
 651 ttccgtgagc cgccatccca agctctggta cacggctgcc caccacacat
 701 cattggatca aagctctcca ccccttagtg gaacatctcc atcctacaag
 751 tacacattac caggagcaca agatgcaaag gatgatttcc cccttcgaaa
 801 aactgaatcc tcagtcagta gcagttctcc aggctctggt cccagttcac
 851 caaacaatgg gccaactgga agtgttactg aaaatgagac ttcggttttg
 901 ccccctaccc ctcatgccga gcaaatggtt tcacagcaac gcattctaat
 951 tcatgaagat tccatgaacc tgctaagtct ttatacctct ccttctttgc
1001 ccaacattac cttggggctt cccgcagtgc catcccagct caatgcttcg
```

FIG. 1H

```
1051  aattcactca aagaaaagca gaagtgtgag acgcagacgc ttaggcaagg
1101  tgttcctctg cctgggcagt atggaggcag catcccggca tcttccagcc
1151  accctcatgt tactttagag ggaaagccac ccaacagcag ccaccaggct
1201  ctcctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct
1251  tgtagctggt ggagttccct tacatcctca gtctcccttg gcaacaaaag
1301  agagaatttc acctggcatt agaggtaccc acaaattgcc ccgtcacaga
1351  cccctgaacc gaacccagtc tgcacctttg cctcagagca cgttggctca
1401  gctggtcatt caacagcaac accagcaatt cttggagaag cagaagcaat
1451  accagcagca gatccacatg aacaaactgc tttcgaaatc tattgaacaa
1501  ctgaagcaac caggcagtca ccttgaggaa gcagaggaag agcttcaggg
1551  ggaccaggcg atgcaggaag acagagcgcc ctctagtggc aacagcacta
1601  ggagcgacag cagtgcttgt gtggatgaca cactgggaca agttggggct
1651  gtgaaggtca aggaggaacc agtggacagt gatgaagatg ctcagatcca
1701  ggaaatggaa tctggggagc aggctgcttt tatgcaacag gtaataggca
1751  aagatttagc tccaggattt gtaattaaag tcattatctg a
```

FIG. 1I

>HDAC9 (deltaNLS)

```
   1 ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca
  51 cttgcaggac tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact
 101 ctaagccaga tggggtggct ggacgagagc agctcttggc tcagcaaaga
 151 atgcacagta tgatcagctc agtggatgtg aagtcagaag ttcctgtggg
 201 cctggagccc atctcacctt tagacctaag gacagacctc aggatgatga
 251 tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta
 301 cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga
 351 gtttcagaaa cagcatgaga acttgacacg gcagcaccag gctcagcttc
 401 aggagcatat caaggaactt ctagccataa aacagcaaca agaactccta
 451 gaaaaggagc agaaactgga gcagcagagg caagaacagg aagtagagag
 501 gcatcgcaga gaacagcagc ttcctcctct cagaggcaaa gatagaggac
 551 gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc
 601 ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca
 651 ttccgtgagc cgccatccca agctctggta cacggctgcc caccacacat
 701 cattggatca aagctctcca cccccttagtg gaacatctcc atcctacaag
 751 tacacattac caggagcaca agatgcaaag gatgatttcc cccttcgaaa
 801 aactgaatcc tcagtcagta gcagttctcc aggctctggt cccagttcac
 851 caaacaatgg gccaactgga agtgttactg aaaatgagac ttcggttttg
 901 ccccctaccc ctcatgccga gcaaatggtt tcacagcaac gcattctaat
 951 tcatgaagat tccatgaacc tgctaagtct ttatacctct ccttctttgc
1001 ccaacattac cttggggctt cccgcagtgc catcccagct caatgcttcg
1051 aattcactca aagaaaagca gaagtgtgag acgcagacgc ttaggcaagg
1101 tgttcctctg cctgggcagt atggaggcag catcccggca tcttccagcc
```

FIG. 1J

1151 accctcatgt tactttagag ggaaagccac ccaacagcag ccaccaggct
1201 ctcctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct
1251 tgtagctggt ggagttccct tacatcctca gtctcccttg gcaacaaaag
1301 agagaatttc acctggcatt agaggtaccc acaaattgcc ccgtcacaga
1351 cccctgaacc gaacccagtc tgcacctttg cctcagagca cgttggctca
1401 gctggtcatt caacagcaac accagcaatt cttggagaag cagaagcaat
1451 accagcagca gatccacatg aacaaactgc tttcgaaatc tattgaacaa
1501 ctgaagcaac caggcagtca ccttgaggaa gcagaggaag agcttcaggg
1551 ggaccaggcg atgcaggaag acagagcgcc ctctagtggc aacagcacta
1601 ggagcgacag cagtgcttgt gtggatgaca cactgggaca agttggggct
1651 gtgaaggtca aggaggaacc agtggacagt gatgaagatg ctcagatcca
1701 ggaaatggaa tctggggagc aggctgcttt tatgcaacag cctttcctgg
1751 aacccacgca cacacgtgcg ctctctgtgc gccaagctcc gctggctgcg
1801 gttggcatgg atggattaga gaaacaccgt ctcgtctcca ggactcactc
1851 ttcccctgct gcctctgttt tacctcaccc agcaatggac cgcccccctcc
1901 agcctggctc tgcaactgga attgcctatg acccccttgat gctgaaacac
1951 cagtgcgttt gtggcaattc caccacccac cctgagcatg ctggacgaat
2001 acagagtatc tggtcacgac tgcaagaaac tgggctgcta aataaatgtg
2051 agcgaattca aggtcgaaaa gccagcctgg aggaaataca gcttgttcat
2101 tctgaacatc actcactgtt gtatggcacc aacccccctgg acggacagaa
2151 gctggacccc aggatactcc taggtgatga ctctcaaaag tttttttcct
2201 cattaccttg tggtggactt ggggtggaca gtgacaccat ttggaatgag
2251 ctacactcgt ccggtgctgc acgcatggct gttggctgtg tcatcgagct
2301 ggcttccaaa gtggcctcag gagagctgaa gaatgggttt gctgttgtga
2351 ggcccccctgg ccatcacgct gaagaatcca cagccatggg gttctgcttt
2401 tttaattcag ttgcaattac cgccaaatac ttgagagacc aactaaatat

FIG. 1K

```
2451   aagcaagata ttgattgtag atctggatgt tcaccatgga aacggtaccc
2501   agcaggcctt ttatgctgac cccagcatcc tgtacatttc actccatcgc
2551   tatgatgaag ggaacttttt ccctggcagt ggagccccaa atgaggttgg
2601   aacaggcctt ggagaagggt acaatataaa tattgcctgg acaggtggcc
2651   ttgatcctcc catgggagat gttgagtacc ttgaagcatt caggaccatc
2701   gtgaagcctg tggccaaaga gtttgatcca gacatggtct tagtatctgc
2751   tggatttgat gcattggaag gccacacccc tcctctagga gggtacaaag
2801   tgacggcaaa atgttttggt catttgacga agcaattgat gacattggct
2851   gatggacgtg tggtgttggc tctagaagga ggacatgatc tcacagccat
2901   ctgtgatgca tcagaagcct gtgtaaatgc ccttctagga aatgagctgg
2951   agccacttgc agaagatatt ctccaccaaa gcccgaatat gaatgctgtt
3001   atttctttac agaagatcat tgaaattcaa agtatgtctt taaagttctc
3051   ttaa
```

FIG. 1L

```
>HDAC9a (deltaNLS)
   1 ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca
  51 cttgcaggac tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact
 101 ctaagccaga tggggtggct ggacgagagc agctcttggc tcagcaaaga
 151 atgcacagta tgatcagctc agtggatgtg aagtcagaag ttcctgtggg
 201 cctggagccc atctcacctt tagacctaag gacagacctc aggatgatga
 251 tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta
 301 cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga
 351 gtttcagaaa cagcatgaga acttgacacg gcagcaccag gctcagcttc
 401 aggagcatat caaggaactt ctagccataa aacagcaaca agaactccta
 451 gaaaaggagc agaaactgga gcagcagagg caagaacagg aagtagagag
 501 gcatcgcaga gaacagcagc ttcctcctct cagaggcaaa gatagaggac
 551 gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc
 601 ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca
 651 ttccgtgagc cgccatccca agctctggta cacggctgcc caccacacat
 701 cattggatca aagctctcca ccccttagtg gaacatctcc atcctacaag
 751 tacacattac caggagcaca agatgcaaag gatgatttcc cccttcgaaa
 801 aactgaatcc tcagtcagta gcagttctcc aggctctggt cccagttcac
 851 caaacaatgg gccaactgga agtgttactg aaaatgagac ttcggttttg
 901 ccccctaccc ctcatgccga gcaaatggtt tcacagcaac gcattctaat
 951 tcatgaagat tccatgaacc tgctaagtct ttatacctct ccttctttgc
1001 ccaacattac cttggggctt cccgcagtgc catcccagct caatgcttcg
1051 aattcactca aagaaaagca gaagtgtgag acgcagacgc ttaggcaagg
1101 tgttcctctg cctgggcagt atggaggcag catcccggca tcttccagcc
1151 accctcatgt tactttagag ggaaagccac ccaacagcag ccaccaggct
```

FIG. 1M

```
1201 ctcctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct
1251 tgtagctggt ggagttccct tacatcctca gtctcccttg gcaacaaaag
1301 agagaatttc acctggcatt agaggtaccc acaaattgcc ccgtcacaga
1351 cccctgaacc gaacccagtc tgcacctttg cctcagagca cgttggctca
1401 gctggtcatt caacagcaac accagcaatt cttggagaag cagaagcaat
1451 accagcagca gatccacatg aacaaactgc tttcgaaatc tattgaacaa
1501 ctgaagcaac caggcagtca ccttgaggaa gcagaggaag agcttcaggg
1551 ggaccaggcg atgcaggaag acagagcgcc ctctagtggc aacagcacta
1601 ggagcgacag cagtgcttgt gtggatgaca cactgggaca agttggggct
1651 gtgaaggtca aggaggaacc agtggacagt gatgaagatg ctcagatcca
1701 ggaaatggaa tctggggagc aggctgcttt tatgcaacag cctttcctgg
1751 aacccacgca cacacgtgcg ctctctgtgc gccaagctcc gctggctgcg
1801 gttggcatgg atggattaga gaaacaccgt ctcgtctcca ggactcactc
1851 ttcccctgct gcctctgttt tacctcaccc agcaatggac cgccccctcc
1901 agcctggctc tgcaactgga attgcctatg accccttgat gctgaaacac
1951 cagtgcgttt gtggcaattc caccacccac cctgagcatg ctggacgaat
2001 acagagtatc tggtcacgac tgcaagaaac tgggctgcta aataaatgtg
2051 agcgaattca aggtcgaaaa gccagcctgg aggaaataca gcttgttcat
2101 tctgaacatc actcactgtt gtatggcacc aaccccctgg acggacagaa
2151 gctggacccc aggatactcc taggtgatga ctctcaaaag tttttttcct
2201 cattaccttg tggtggactt ggggtggaca gtgacaccat ttggaatgag
2251 ctacactcgt ccggtgctgc acgcatggct gttggctgtg tcatcgagct
2301 ggcttccaaa gtggcctcag gagagctgaa gaatgggttt gctgttgtga
2351 ggccccctgg ccatcacgct gaagaatcca cagccatggg gttctgcttt
2401 tttaattcag ttgcaattac cgccaaatac ttgagagacc aactaaatat
```

FIG. 1N

```
2451  aagcaagata ttgattgtag atctggatgt tcaccatgga aacggtaccc
2501  agcaggcctt ttatgctgac cccagcatcc tgtacatttc actccatcgc
2551  tatgatgaag ggaacttttt ccctggcagt ggagccccaa atgaggttcg
2601  gtttatttct ttagagcccc acttttattt gtatctttca ggtaattgca
2651  ttgcatgatt accccctaatt ttcttgtcct ttgctggtgt tttaaattac
2701  acgagattac tgaattgtcc catgggacca agaaccagtg cagaacaagt
2751  gcataaccca gagcactgtt tgtcagggaa ggttgggctg atttgatgtg
2801  ttgtttgatg tttatttcaa gagctcccat gtgcttgttt tcctctcttc
2851  ttgctttctt ccatttgctc tcttctctgc ccaccgtggt gtgtctttct
2901  cttcccaggt tggaacaggc cttggagaag ggtacaatat aaatattgcc
2951  tggacaggtg gccttgatcc tcccatggga gatgttgagt accttgaagc
3001  attcaggacc atcgtgaagc ctgtggccaa agagtttgat ccagacatgg
3051  tcttagtatc tgctggattt gatgcattgg aaggccacac ccctcctcta
3101  ggagggtaca aagtgacggc aaaatgtttt ggtcatttga cgaagcaatt
3151  gatgacattg gctgatggac gtgtggtgtt ggctctagaa ggaggacatg
3201  atctcacagc catctgtgat gcatcagaag cctgtgtaaa tgcccttcta
3251  ggaaatgagc tggagccact tgcagaagat attctccacc aaagcccgaa
3301  tatgaatgct gttatttctt tacagaagat cattgaaatt caaagtatgt
3351  ctttaaagtt ctcttaa
```

FIG. 1O

FIG. 2A
FIG. 2B
FIG. 2C
FIG. 2D
FIG. 2E

FIG. 2

```
>HDAC9 (1011 amino acids)
MHSMISSVDVKSEVPVGLEPISPLDLRTDLRMMMPVVDPVVREKQLQQELLLIQQQQQI
QKQLLIAEFQKQHENLTRQHQAQLQEHIKELLAIKQQQELLEKEQKLEQQRQEQEVERH
RREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDTPTNGKNHSVSRHPKLWY
TAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKDDFPLRKTASEPNLKVRSRLKQKVAE
RRSSPLLRRKDGNVVTSFKKRMFEVTESSVSSSSPGSGPSSPNNGPTGSVTENETSVLP
PTPHAEQMVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQLNASNSLKEKQKCE
TQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALLQHLLLKEQMRQQKLLVA
GGVPLHPQSPLATKERISPGIRGTHKLPRHRPLNRTQSAPLPQSTLAQLVIQQQHQQFL
EKQKQYQQQIHMNKLLSKSIEQLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDS
SACVDDTLGQVGAVKVKEEPVDSDEDAQIQEMESGEQAAFMQQPFLEPTHTRALSVRQA
PLAAVGMDGLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGSATGIAYDPLMLKHQCVCG
NSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKASLEEIQLVHSEHHSLLYGTNPLD
GQKLDPRILLGDDSQKFFSSLPCGGLGVDSDTIWNELHSSGAARMAVGCVIELASKVAS
GELKNGFAVVRPPGHHAEESTAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNG
TQQAFYADPSILYISLHRYDEGNFFPGSGAPNEVGTGLGEGYNINIAWTGGLDPPMGDV
EYLEAFRTIVKPVAKEFDPDMVLVSAGFDALEGHTPPLGGYKVTAKCFGHLTKQLMTLA
DGRVVLALEGGHDLTAICDASEACVNALLGNELEPLAEDILHQSPNMNAVISLQKIIEI
QSMSLKFS
```

FIG. 2A

```
>HDAC9a (879 amino acids)
MHSMISSVDVKSEVPVGLEPISPLDLRTDLRMMMPVVDPVVREKQLQQELLLIQQQQQI
QKQLLIAEFQKQHENLTRQHQAQLQEHIKELLAIKQQQELLEKEQKLEQQRQEQEVERH
RREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDTPTNGKNHSVSRHPKLWY
TAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKDDFPLRKTASEPNLKVRSRLKQKVAE
RRSSPLLRRKDGNVVTSFKKRMFEVTESSVSSSSPGSGPSSPNNGPTGSVTENETSVLP
PTPHAEQMVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQLNASNSLKEKQKCE
TQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALLQHLLLKEQMRQQKLLVA
GGVPLHPQSPLATKERISPGIRGTHKLPRHRPLNRTQSAPLPQSTLAQLVIQQQHQQFL
EKQKQYQQQIHMNKLLSKSIEQLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDS
SACVDDTLGQVGAVKVKEEPVDSDEDAQIQEMESGEQAAFMQQPFLEPTHTRALSVRQA
PLAAVGMDGLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGSATGIAYDPLMLKHQCVCG
NSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKASLEEIQLVHSEHHSLLYGTNPLD
GQKLDPRILLGDDSQKFFSSLPCGGLGVDSDTIWNELHSSGAARMAVGCVIELASKVAS
GELKNGFAVVRPPGHHAEESTAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNG
TQQAFYADPSILYISLHRYDEGNFFPGSGAPNEVRFISLEPHFYLYLSGNCIA
```

FIG. 2B

>HDAC9 (ΔNLS) (967 amino acids)
MHSMISSVDVKSEVPVGLEPISPLDLRTDLRMMMPVVDPVVREKQLQQELLLIQQQQQI
QKQLLIAEFQKQHENLTRQHQAQLQEHIKELLAIKQQQELLEKEQKLEQQRQEQEVERH
RREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDTPTNGKNHSVSRHPKLWY
TAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKDDFPLRKTESSVSSSSPGSGPSSPNN
GPTGSVTENETSVLPPTPHAEQMVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPS
QLNASNSLKEKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALLQH
LLLKEQMRQQKLLVAGGVPLHPQSPLATKERISPGIRGTHKLPRHRPLNRTQSAPLPQS
TLAQLVIQQQHQQFLEKQKQYQQQIHMNKLLSKSIEQLKQPGSHLEEAEEELQGDQAMQ
EDRAPSSGNSTRSDSSACVDDTLGQVGAVKVKEEPVDSDEDAQIQEMESGEQAAFMQQP
FLEPTHTRALSVRQAPLAAVGMDGLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGSATG
IAYDPLMLKHQCVCGNSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKASLEEIQLV
HSEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLPCGGLGVDSDTIWNELHSSGAAR
MAVGCVIELASKVASGELKNGFAVVRPPGHHAEESTAMGFCFFNSVAITAKYLRDQLNI
SKILIVDLDVHHGNGTQQAFYADPSILYISLHRYDEGNFFPGSGAPNEVGTGLGEGYNI
NIAWTGGLDPPMGDVEYLEAFRTIVKPVAKEFDPDMVLVSAGFDALEGHTPPLGGYKVT
AKCFGHLTKQLMTLADGRVVLALEGGHDLTAICDASEACVNALLGNELEPLAEDILHQS
PNMNAVISLQKIIEIQSMSLKFS

FIG. 2C

>HDAC9a (ΔNLS) (835 amino acids)
MHSMISSVDVKSEVPVGLEPISPLDLRTDLRMMMPVVDPVVREKQLQQELLLIQQQQQI
QKQLLIAEFQKQHENLTRQHQAQLQEHIKELLAIKQQQELLEKEQKLEQQRQEQEVERH
RREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDTPTNGKNHSVSRHPKLWY
TAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKDDFPLRKTESSVSSSSPGSGPSSPNN
GPTGSVTENETSVLPPTPHAEQMVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPS
QLNASNSLKEKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALLQH
LLLKEQMRQQKLLVAGGVPLHPQSPLATKERISPGIRGTHKLPRHRPLNRTQSAPLPQS
TLAQLVIQQQHQQFLEKQKQYQQQIHMNKLLSKSIEQLKQPGSHLEEAEEELQGDQAMQ
EDRAPSSGNSTRSDSSACVDDTLGQVGAVKVKEEPVDSDEDAQIQEMESGEQAAFMQQP
FLEPTHTRALSVRQAPLAAVGMDGLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGSATG
IAYDPLMLKHQCVCGNSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKASLEEIQLV
HSEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLPCGGLGVDSDTIWNELHSSGAAR
MAVGCVIELASKVASGELKNGFAVVRPPGHHAEESTAMGFCFFNSVAITAKYLRDQLNI
SKILIVDLDVHHGNGTQQAFYADPSILYISLHRYDEGNFFPGSGAPNEVRFISLEPHFY
LYLSGNCIA

FIG. 2D

>HDRPa (HDRP ΔNLS)(546 amino acids)
MHSMISSVDVKSEVPVGLEPISPLDLRTDLRMMMPVVDPVVREKQLQQELLLIQQQQQI
QKQLLIAEFQKQHENLTRQHQAQLQEHIKELLAIKQQQELLEKEQKLEQQRQEQEVERH
RREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDTPTNGKNHSVSRHPKLWY
TAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKDDFPLRKTESSVSSSSPGSGPSSPNN
GPTGSVTENETSVLPPTPHAEQMVSQQRILIHEDSMNLLSLYTSPSLPNITLGLPAVPS
QLNASNSLKEKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSHQALLQH
LLLKEQMRQQKLLVAGGVPLHPQSPLATKERISPGIRGTHKLPRHRPLNRTQSAPLPQS
TLAQLVIQQQHQQFLEKQKQYQQQIHMNKLLSKSIEQLKQPGSHLEEAEEELQGDQAMQ
EDRAPSSGNSTRSDSSACVDDTLGQVGAVKVKEEPVDSDEDAQIQEMESGEQAAFMQQV
IGKDLAPGFVIKVII

FIG. 2E

FIG. 3A
FIG. 3B
FIG. 3C

FIG. 3

FIG. 3A

```
HDRP      1  ------------------------MHSMISSVDVKSEVPVGLEP-ISPLDLRTDLRMMMP
HDAC9a    1  ------------------------MHSMISSVDVKSEVPVGLEP-ISPLDLRTDLRMMMP
HDAC9     1  ------------------------MHSMISSVDVKSEVPVGLEP-ISPLDLRTDLRMMMP
HDAC4     1  MSSQSHPDGLSGRDQPVELLNPARVNHMPSTVDVATALPLQVAPSAVPMDLRLDHQFSLP

HDRP     36  VVDPVVREKQLQQELLLIQQQQQIQKQLLIAEFQKQHENLTRQHQAQLQEHIK---ELLA
HDAC9a   36  VVDPVVREKQLQQELLLIQQQQQIQKQLLIAEFQKQHENLTRQHQAQLQEHIK---ELLA
HDAC9    36  VVDPVVREKQLQQELLLIQQQQQIQKQLLIAEFQKQHENLTRQHQAQLQEHIK---ELLA
HDAC4    61  VAEPALREQQLQQELLALKQKQQIQRQILIAEFQRQHEQLSRQHEAQLHEHIKQQQEMLA

HDRP     93  IKQQQELLEKEQKLEQQRQEQEVERHRREQQLPPLRGKDRGRERAVASTEVKQKLQEFLL
HDAC9a   93  IKQQQELLEKEQKLEQQRQEQEVERHRREQQLPPLRGKDRGRERAVASTEVKQKLQEFLL
HDAC9    93  IKQQQELLEKEQKLEQQRQEQEVERHRREQQLPPLRGKDRGRERAVASTEVKQKLQEFLL
HDAC4   121  MKHQQELLEHQRKLERHRQEQELEKQHREQKLQQLKNKEKGKESAVASTEVKMKLQEFVL

HDRP    153  SKSATKDTPTNGKNHSVSRHPKLWYTAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKDD
HDAC9a  153  SKSATKDTPTNGKNHSVSRHPKLWYTAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKDD
HDAC9   153  SKSATKDTPTNGKNHSVSRHPKLWYTAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKDD
HDAC4   181  NK--KKALAHRNLNHCISSDPRYWYGKTQHSSLDQSSPPQSGVSTSYNHPVLGMYDAKDD

HDRP    213  FPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGNVVTSFKKRMFEVTESSVSSSSPG
HDAC9a  213  FPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGNVVTSFKKRMFEVTESSVSSSSPG
HDAC9   213  FPLRKTASEPNLKVRSRLKQKVAERRSSPLLRRKDGNVVTSFKKRMFEVTESSVSSSSPG
HDAC4   239  FPLRKTASEPNLKLRSRLKQKVAERRSSPLLRRKDGPVVTALKKRPLDVTDSACSS-APG
```

```
HDRP     273  SGPSSPNNGPTGSVTENETSVLPPTPHAEQMVSQQRILIHEDSMNLLSLYTSPSLPNITL
HDAC9a   273  SGPSSPNNGPTGSVTENETSVLPPTPHAEQMVSQQRILIHEDSMNLLSLYTSPSLPNITL
HDAC9    273  SGPSSPNNGPTGSVTENETSVLPPTPHAEQMVSQQRILIHEDSMNLLSLYTSPSLPNITL
HDAC4    298  SGPSSPNNSSGSVSAENGIAPAVPSTPAETSLAHR-LVAREGSAAPLPLYTSPSLPNITL

HDRP     333  GLPAVPSQLNASNSLKEKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSH
HDAC9a   333  GLPAVPSQLNASNSLKEKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSH
HDAC9    333  GLPAVPSQLNASNSLKEKQKCETQTLRQGVPLPGQYGGSIPASSSHPHVTLEGKPPNSSH
HDAC4    357  GLPATGPSAGTAGQQ-DTERLTLPALQQRLSLFPGTHLTPYLSTS--PLERDG---GAAH

HDRP     393  QALLQHLLLKEQMRQQKLLVAGG--VPLHPQSPLATKERISPGIRGTHKLPRHRPLNRTQ
HDAC9a   393  QALLQHLLLKEQMRQQKLLVAGG--VPLHPQSPLATKERISPGIRGTHKLPRHRPLNRTQ
HDAC9    393  QALLQHLLLKEQMRQQKLLVAGG--VPLHPQSPLATKERISPGIRGTHKLPRHRPLNRTQ
HDAC4    411  SPLLQHMVLLEQPPAQAPLVTGLGALPLHAQS-LVGADRVSP---SIHKLRQHRPLGRTQ

HDRP     451  SAPLPQ--STLAQLVIQQQHQQFLEKQKQ--YQQQIHMNKLLSKSIEQLKQPGSHLEEAE
HDAC9a   451  SAPLPQ--STLAQLVIQQQHQQFLEKQKQ--YQQQIHMNKLLSKSIEQLKQPGSHLEEAE
HDAC9    451  SAPLPQ--STLAQLVIQQQHQQFLEKQKQ--YQQQIHMNKLLSKSIEQLKQPGSHLEEAE
HDAC4    467  SAPLPQNAQALQHLVIQQQHQQFLEKHKQQFQQQQLQMNKIIPKPSEPARQPESHPEETE

HDRP     507  EELQGDQAMQEDRAPSSGNSTR-SDSSACVDDTLGQVGAVKVKEEPVDSDEDAQIQEMES
HDAC9a   507  EELQGDQAMQEDRAPSSGNSTR-SDSSACVDDTLGQVGAVKVKEEPVDSDEDAQIQEMES
HDAC9    507  EELQGDQAMQEDRAPSSGNSTR-SDSSACVDDTLGQVGAVKVKEEPVDSDEDAQIQEMES
HDAC4    527  EELREHQALLDEPYLDRLPGQKEAHAQAGVQVKQEPIESDEEEAEPPREVEPGQRQPSEQ

HDRP     566  GEQAAFMQQVIGKDLAPGFVIKVII-----------------------------------
HDAC9a   566  GEQAAFMQQPFLEPTHTRALSVRQAPLAAVGMDGLEKHRLVSRTHSSPAASVLPHPAMDR
HDAC9    566  GEQAAFMQQPFLEPTHTRALSVRQAPLAAVGMDGLEKHRLVSRTHSSPAASVLPHPAMDR
HDAC4    587  ELLFRQQALLLEQQRIHQLRNYQASMEAAGIPVSFGGHRPLSRAQSSPASATFPVSVQEP
```

FIG. 3B

```
HDRP          ------------------------------------------------------------
HDAC9a   626  PLQPGSATGIAYDPLMLKHQCVCGNSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKA
HDAC9    626  PLQPGSATGIAYDPLMLKHQCVCGNSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKA
HDAC4    647  PTKPRFTTGLVYDTLMLKHQCTCGSSSSHPEHAGRIQSIWSRLQETGLRGKCECIRGRKA

HDRP          ------------------------------------------------------------
HDAC9a   686  SLEEIQLVHSEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLPCGGLGVDSDTIWNEL
HDAC9    686  SLEEIQLVHSEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLPCGGLGVDSDTIWNEL
HDAC4    707  TLEELQTVHSEAHTLLYGTNPLNRQKLDSKKLLGSLASVFVR-LPCGGVGVDSDTIWNEV

HDRP          ------------------------------------------------------------
HDAC9a   746  HSSGAARMAVGCVIELASKVASGELKNGFAVVRPPGHHAEESTAMGFCFFNSVAITAKYL
HDAC9    746  HSSGAARMAVGCVIELASKVASGELKNGFAVVRPPGHHAEESTAMGFCFFNSVAITAKYL
HDAC4    766  HSAGAARLAVGCVVELVFKVATGELKNGFAVVRPPGHHAEESTPMGFCYFNSVAVAAKLL

HDRP          ------------------------------------------------------------
HDAC9a   806  RDQLNISKILIVDLDVHHGNGTQQAFYADPSILYISLHRYDEGNFFPGSGAPNEV RFISL
HDAC9    806  RDQLNISKILIVDLDVHHGNGTQQAFYADPSILYISLHRYDEGNFFPGSGAPNEVGTG LG
HDAC4    826  QQRLSVSKILIVDWDVHHGNGTQQAFYSDPSVLYMSLHRYDDGNFFPGSGAPDEVGTGPG

HDRP          ------------------------------------------------------------
HDAC9a   866  EPHFYLYLSGNCIA----------------------------------------------
HDAC9    866  EGYNINIAWTGGLDPPMGDVEYLEAFRTIVKPVAKEFDPDMVLVSAGFDALEGHTPPLGG
HDAC4    886  VGFNVNMAFTGGLDPPMGDAEYLAAFRTVVMPIASEFAPDVVLVSSGFDAVEGHPTPLGG

HDRP          ------------------------------------------------------------
HDAC9a        ------------------------------------------------------------
HDAC9    926  YKVTAKCFGHLTKQLMTLADGRVVLALEGGHDLTAICDASEACVNALLGNELEPLAEDIL
HDAC4    946  YNLSARCFGYLTKQLMGLAGGRIVLALEGGHDLTAICDASEACVSALLGNELDPLPEKVL

HDRP          ------------------------------------------------------------
HDAC9a        ------------------------------------------------------------
HDAC9    986  HQSPNMNAVISLQKIIEIQSMSLKFS----------------------------------
HDAC4   1006  QQRPNANAVRSMEKVMEIHSKYWRCLQRTTSTAGRSLIEAQTCENEEAETVTAMASLSVG

HDRP          --------------------
HDAC9a        --------------------
HDAC9         --------------------
HDAC4   1066  VKPAEKRPDEEPMEEEPPL
```

FIG. 3C

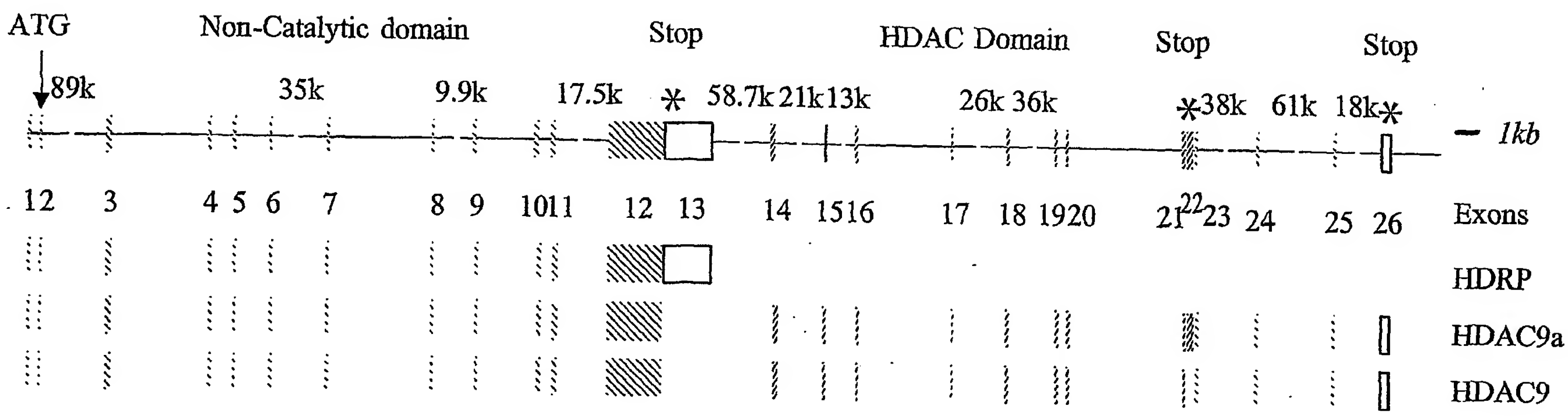


FIG. 4

FIG. 5A
FIG. 5B
FIG. 5C
FIG. 5D

FIG. 5

1 /[1]ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact

101 ctaagccag/[2]a tggggtggct ggacgagagc agctcttggc tcagcaaaga **ATG**CACAGTA TGATCAGCTC AGT/[3]GGATGTG AAGTCAGAAG TTCCTGTGGG

201 CCTGGAGCCC ATCTCACCTT TAGACCTAAG GACAGACCTC AGGATGATGA TGCCCGTGGT GGACCCTGTT GTCCGTGAGA AGCAATTGCA GCAGGAATTA

301 CTTCTTATCC AGCAGCAGCA ACAAATCCAG AAGCAGCTTC TGATAGCAGA GTTTCAGAAA CAGCATGAGA ACTTGACACG GCAGCACCAG GCTCAGCTTC

401 AGGAGCATAT CAAG/[4]GAACTT CTAGCCATAA AACAGCAACA AGAACTCCTA GAAAAGGAGC AGAAACTGGA GCAGCAGAGG CAAGAACAGG AAGTAGAGAG

501 GCATCGCAGA GAACAGCAGC TTCCTCCTCT CAGAGGCAAA GATAGAGGAC GAGAAAG /[5]GGC AGTGGCAAGT ACAGAAGTAA AGCAGAAGCT TCAAGAGTTC

601 CTACTGAGTA AATCAGCAAC GAAAGACACT CCAACTAATG GAAAAAATCA TTCCGTGAGC CGCCATCCCA AGCTCTGGTA CACG/[6]GCTGCC CACCACACAT

701 CATTGGATCA AAGCTCTCCA CCCCTTAGTG GAACATCTCC ATCCTACAAG TACACATTAC CAGGAGCACA AGATGCAAAG GATGATTTCC CCCTTCGAAA

FIG. 5A

801 AACT/[7]GCCTCT GAGCCCAACT TGAAGGTGCG GTCCAGGTTA AAACAGAAAG TGGCAGAGAG GAGAAGCAGC CCCTTACTCA GGCGGAAGGA TGGAAATGTT

901 GTCACTTCAT TCAAGAAGCG AATGTTTGAG GTGACAG/[8]AAT CCTCAGTCAG TAGCAGTTCT CCAGGCTCTG GTCCCAGTTC ACCAAACAAT GGGCCAACTG

1001 GAAGTGTTAC TGAAAATGAG ACTTCGGTTT TGCCCCCTAC CCCTCATGCC GAG /[9]CAAATGG TTTCACAGCA ACGCATTCTA ATTCATGAAG ATTCCATGAA

1101 CCTGCTAAGT CTTTATACCT CTCCTTCTTT GCCCAACATT ACCTTGGGGC TTCCCGCAGT GCCATCCCAG CTCAATG /[10]CTT CGAATTCACT CAAAGAAAAG

1201 CAGAAGTGTG AGACGCAGAC GCTTAGGCAA GGTGTTCCTC TGCCTGGGCA GTATGGAGGC AGCATCCCGG CATCTTCCAG CCACCCTCAT GTTACTTTAG

1301 AGGGAAAGCC ACCCAACAGC AGCCACCAGG CTCTCCTGCA GCATTTATTA TTGAAAGAAC AAATGCGACA GCAAAAGCTT CTTGTAGCTG/[11] GTGGAGTTCC

1401 CTTACATCCT CAGTCTCCCT TGGCAACAAA AGAGAGAATT TCACCTGGCA TTAGAGGTAC CCACAAATTG CCCCGTCACA GACCCCTGAA CCGAACCCAG

1501 TCTGCACCTT TGCCTCAGAG CACGTTGGCT CAGCTGGTCA TTCAACAGCA ACACCAGCAA TTCTTGGAGA AGCAGAAGCA ATACCAGCAG CAGATCCACA

1601 TGAACAAA/[12]CT GCTTTCGAAA TCTATTGAAC AACTGAAGCA ACCAGGCAGT CACCTTGAGG AAGCAGAGGA AGAGCTTCAG GGGGACCAGG CGATGCAGGA

FIG. 5B

1701 AGACAGAGCG CCCTCTAGTG GCAACAGCAC TAGGAGCGAC AGCAGTGCTT GTGTGGATGA CACACTGGGA CAAGTTGGGG CTGTGAAGGT CAAGGAGGAA

1801 CCAGTGGACA GTGATGAAGA TGCTCAGATC CAGGAAATGG AATCTGGGGA GCAGGCTGCT TTTATGCAAC AG

/[13]GTAATAGG CAAAGATTTA GCTCCAGGAT TTGTAATTAA AGTCATTATC TGA..... /[14]CCTTTCCT GGAACCCACG CACACACGTG

1901 CGCTCTCTGT GCGCCAAGCT CCGCTGGCTG CGGTTGGCAT GGATGGATTA GAGAAACACC GTCTCGTCTC CAGGACTCAC TCTTCCCCTG CTGCCTCTGT

2001 TTTACCTCAC CCAGCAATGG ACCGCCCCCT CCAGCCTGGC TCTGCAACTG /[15]GAATTGCCTA TGACCCCTTG ATGCTGAAAC ACCAGTGCGT TTGTGGCAAT

2101 TCCACCACCC ACCCTGAGCA TGCTGGACGA ATACAGAGTA TCTGGTCACG ACTGCAAGAA ACTGGGCTGC TAAATAAATG TGAG/[16]CGAATT CAAGGTCGAA

2201 AAGCCAGCCT GGAGGAAATA CAGCTTGTTC ATTCTGAACA TCACTCACTG TTGTATGGCA CCAACCCCCT GGACGGACAG AAGCTGGACC CCAGGATACT

2301 CCTAG/[17]GTGAT GACTCTCAAA AGTTTTTTTC CTCATTACCT TGTGGTGGAC TTGGG/[18]GTGGA CAGTGACACC ATTTGGAATG AGCTACACTC GTCCGGTGCT

2401 GCACGCATGG CTGTTGGCTG TGTCATCGAG CTGGCTTCCA AAGTGGCCTC AGGAGAGCTG AAGA /[19]ATGGGT TTGCTGTTGT GAGGCCCCCT GGCCATCACG

2501 CTGAAGAATC CACAGCCATG /[20]GGGTTCTGCT TTTTTAATTC AGTTGCAATT ACCGCCAAAT ACTTGAGAGA CCAACTAAAT ATAAGCAAGA TATTGATTGT

FIG. 5C

2601 AGATCTG/[21]GAT GTTCACCATG GAAACGGTAC CCAGCAGGCC TTTTATGCTG ACCCCAGCAT CCTGTACATT TCACTCCATC GCTATGATGA AGGGAACTTT

2701 TTCCCTGGCA GTGGAGCCCC AAATGAGG/[22]TT CGGTTTATTT CTTTAGAGCC CCACTTTTAT TTGTATCTTT CAGGTAATTG CATTGCA**TGA** ttacccctaa

2801 ttttcttgtc ctttgctggt gttttaaatt acacgagatt actgaattgt cccatgggac caagaaccag tgcagaacaa gtgcataacc cagagcactg

2901 tttgtcaggg aaggttgggc tgatttgatg tgttgtttga tgtttatttc aagagctccc atgtgcttgt tttcctctct tcttgctttc ttccatttgc

3001 tctcttctct gcccaccgtg gtgtgtcttt ctcttcccag /[23]gttggaacag gccttggaga agggtacaat ataaatattg cctggacagg tggccttgat

3101 cctcccatgg gagatgttga gtaccttgaa gcattcag/[24]ga ccatcgtgaa gcctgtggcc aaagagtttg atccagacat ggtcttagta tctgctggat

3201 ttgatgcatt ggaaggccac acccctcctc taggagggta caaagtgacg gcaaaatg/[25]tt ttggtcattt gacgaagcaa ttgatgacat tggctgatgg

3301 acgtgtggtg ttggctctag aaggaggaca tgatctcaca gccatctgtg atgcatcaga agcctgtgta aatgcccttc taggaaatga g/[26]ctggagcca

3401 cttgcagaag atattctcca ccaaagcccg aatatgaatg ctgttatttc tttacagaag atcattgaaa ttcaaagtat gtctttaaag ttctct**taa**.....

FIG. 5D

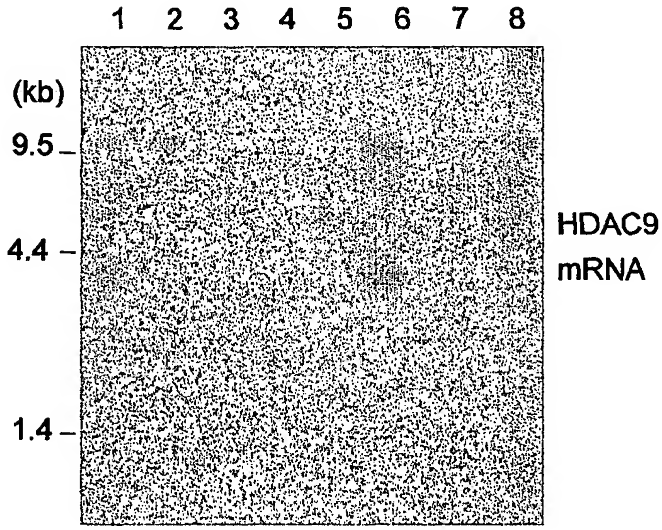


FIG. 6A

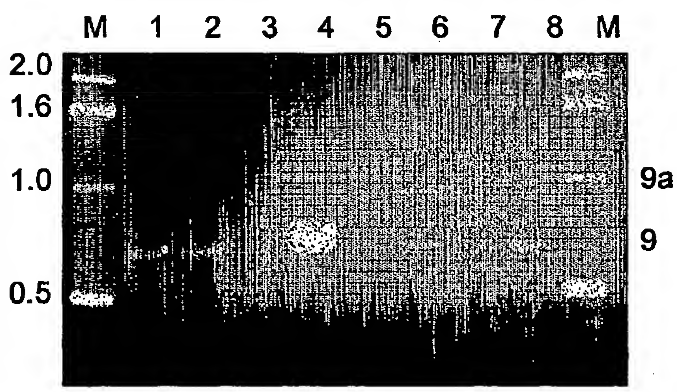


FIG. 6B

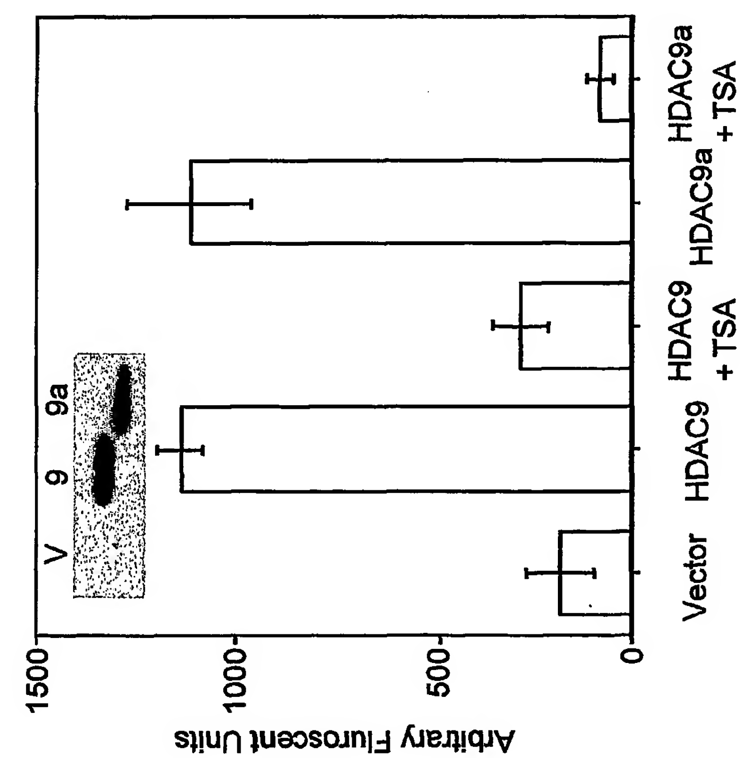

Arbitrary Fluroscent Units
1500
1000
500
0
V
9
9a
Vector
HDAC9
HDAC9
+TSA
HDAC9a
HDAC9a
+TSA

FIG. 7

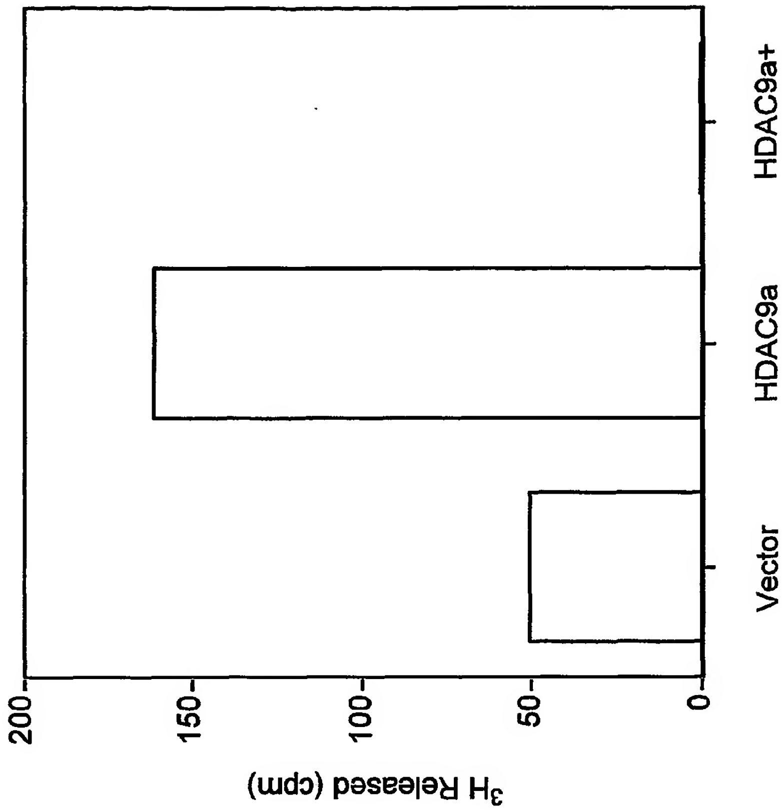

^{3}H Released (cpm)
200
150
100
50
0
Vector
HDAC9a
HDAC9a+

FIG. 8

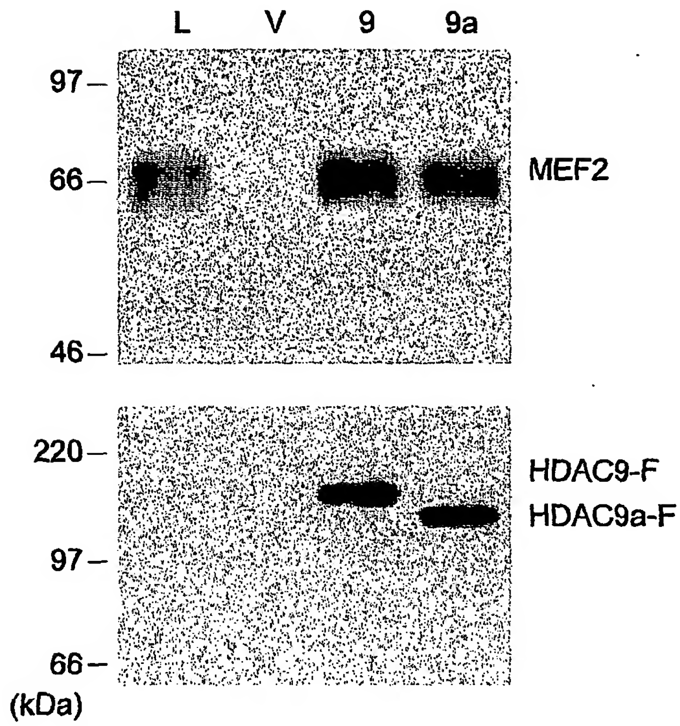

L
V
9
9a
97
66
46
MEF2
220
97
66
(kDa)
HDAC9-F
HDAC9a-F

FIG. 9A

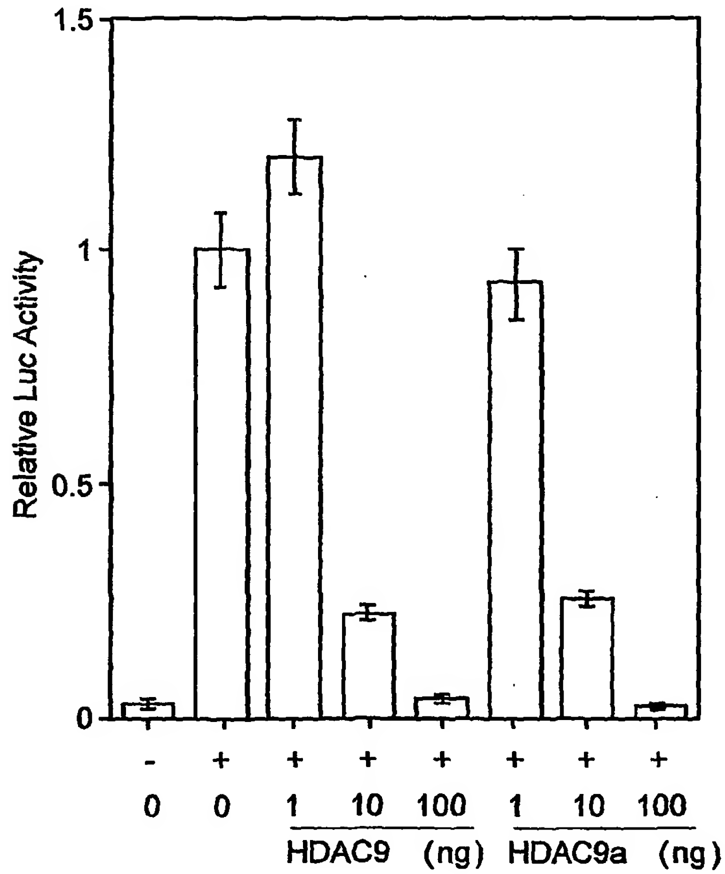

Relative Luc Activity
1.5
1
0.5
0
- + + + + + + +
0 0 1 10 100 1 10 100
HDAC9 (ng) HDAC9a (ng)

FIG. 9B

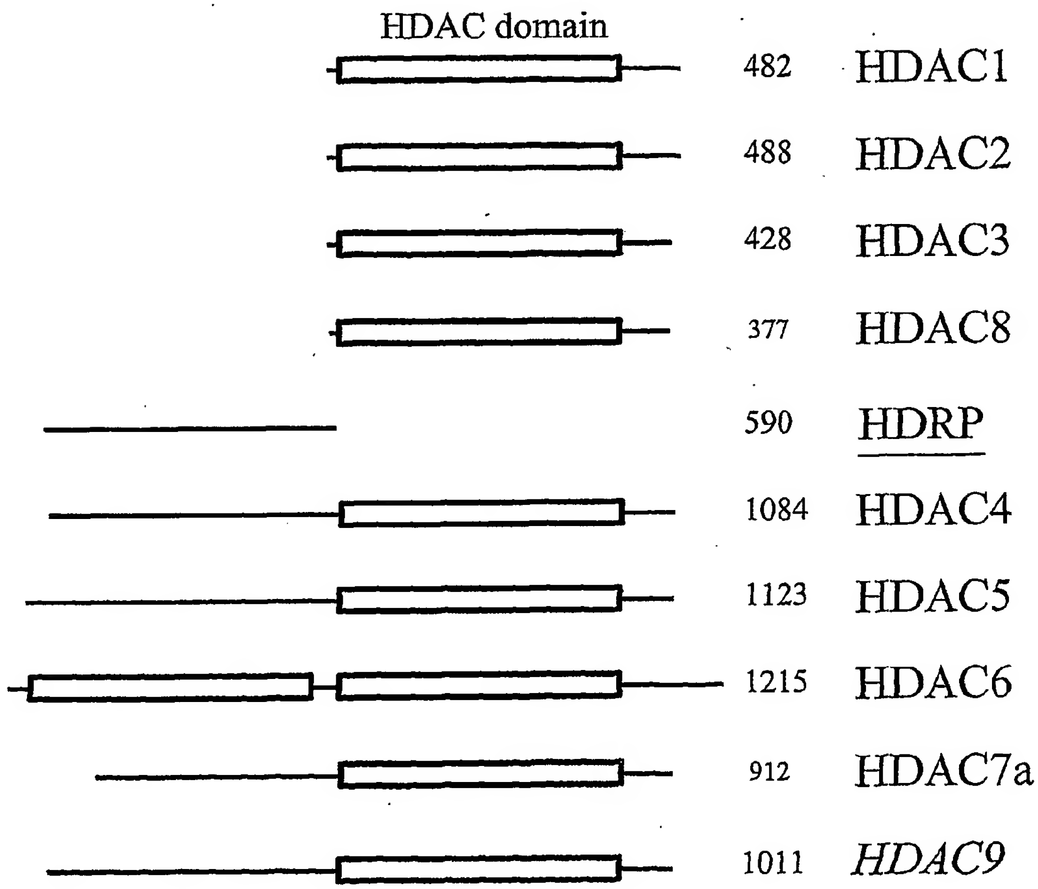

HDAC domain
482 HDAC1
488 HDAC2
428 HDAC3
377 HDAC8
590 HDRP
1084 HDAC4
1123 HDAC5
1215 HDAC6
912 HDAC7a
1011 HDAC9

FIG. 10

FIG. 11A
FIG. 11B
FIG. 11C
FIG. 11D
FIG. 11E
FIG. 11F

FIG. 11

FIG. 11A

cccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggg
ggatgtgctgcaaggcgattaagttgggtaacgcccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagct
gatctaatcaatattggccattagccatattattcattggttatatagcataaatcaatattggctattggccattgcatacgttgtatcca
tatcataatatgtacatttatattggctcatgtccaacattaccgccatgttgacattgattattgactagttattaatagtaatcaattacg
gggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggcgaccgcccagcgaccc
ccgcccgttgacgtcaatagtgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacg

gtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtccgccccctattgacgtcaatgacggtaaatggcccgcct
agcattatgcccagtacatgaccttacgggagtttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcg
gttttggcagtacaccaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtt
tgttttggcaccaaaatcaacgggactttccaaaatgtcgtaataaccccgccccgttgacgcaaatgggcggtaggcgtgtacg
gtgggaggtctatataagcagagctcgtttagtgaaccgtcagaattcaagcttgcggccgcagatctatcgatctgcaggatatc
(EcoRV)

acc

ATGCACAGTATGATCAGCTCAGTGGATGTGAAGTCAGAAGTTCCTGTGGG
CCTGGAGCCCATCTCACCTTTAGACCTAAGGACAGACCTCAGGATGATGA
TGCCCGTGGTGGACCCTGTTGTCCGTGAGAAGCAATTGCAGCAGGAATTA
CTTCTTATCCAGCAGCAGCAACAAATCCAGAAGCAGCTTCTGATAGCAGA
GTTTCAGAAACAGCATGAGAACTTGACACGGCAGCACCAGGCTCAGCTTC
AGGAGCATATCAAGGAACTTCTAGCCATAAAACAGCAACAAGAACTCCTA
GAAAAGGAGCAGAAACTGGAGCAGCAGAGGCAAGAACAGGAAGTAGAGAG
GCATCGCAGAGAACAGCAGCTTCCTCCTCTCAGAGGCAAAGATAGAGGAC
GAGAAAGGGCAGTGGCAAGTACAGAAGTAAAGCAGAAGCTTCAAGAGTTC
CTACTGAGTAAATCAGCAACGAAAGACACTCCAACTAATGGAAAAAATCA
TTCCGTGAGCCGCCATCCCAAGCTCTGGTACACGGCTGCCCACCACACAT
CATTGGATCAAAGCTCTCCACCCCTTAGTGGAACATCTCCATCCTACAAG

FIG. 11B

TACACATTACCAGGAGCACAAGATGCAAAGGATGATTTCCCCCTTCGAAA
AACTGCCTCTGAGCCCAACTTGAAGGTGCGGTCCAGGTTAAAACAGAAAG
TGGCAGAGAGGAGAAGCAGCCCCTTACTCAGGCGGAAGGATGGAAATGTT
GTCACTTCATTCAAGAAGCGAATGTTTGAGGTGACAGAATCCTCAGTCAG
TAGCAGTTCTCCAGGCTCTGGTCCCAGTTCACCAAACAATGGGCCAACTG
GAAGTGTTACTGAAAATGAGACTTCGGTTTTGCCCCCTACCCCTCATGCC
GAGCAAATGGTTTCACAGCAACGCATTCTAATTCATGAAGATTCCATGAA
CCTGCTAAGTCTTTATACCTCTCCTTCTTTGCCCAACATTACCTTGGGGC
TTCCCGCAGTGCCATCCCAGCTCAATGCTTCGAATTCACTCAAAGAAAAG
CAGAAGTGTGAGACGCAGACGCTTAGGCAAGGTGTTCCTCTGCCTGGGCA
GTATGGAGGCAGCATCCCGGCATCTTCCAGCCACCCTCATGTTACTTTAG
AGGGAAAGCCACCCAACAGCAGCCACCAGGCTCTCCTGCAGCATTTATTA
TTGAAAGAACAAATGCGACAGCAAAAGCTTCTTGTAGCTGGTGGAGTTCC
CTTACATCCTCAGTCTCCCTTGGCAACAAAAGAGAGAATTTCACCTGGCA
TTAGAGGTACCCACAAATTGCCCCGTCACAGACCCCTGAACCGAACCCAG
TCTGCACCTTTGCCTCAGAGCACGTTGGCTCAGCTGGTCATTCAACAGCA
ACACCAGCAATTCTTGGAGAAGCAGAAGCAATACCAGCAGCAGATCCACA
TGAACAAACTGCTTTCGAAATCTATTGAACAACTGAAGCAACCAGGCAGT
CACCTTGAGGAAGCAGAGGAAGAGCTTCAGGGGGACCAGGCGATGCAGGA
AGACAGAGCGCCCTCTAGTGGCAACAGCACTAGGAGCGACAGCAGTGCTT
GTGTGGATGACACACTGGGACAAGTTGGGGCTGTGAAGGTCAAGGAGGAA
CCAGTGGACAGTGATGAAGATGCTCAGATCCAGGAAATGGAATCTGGGGA
GCAGGCTGCTTTTATGCAACAGCCTTTCCTGGAACCCACGCACACACGTG
CGCTCTCTGTGCGCCAAGCTCCGCTGGCTGCGGTTGGCATGGATGGATTA

FIG. 11C

GAGAAACACCGTCTCGTCTCCAGGACTCACTCTTCCCCTGCTGCCTCTGT
TTTACCTCACCCAGCAATGGACCGCCCCCTCCAGCCTGGCTCTGCAACTG
GAATTGCCTATGACCCCTTGATGCTGAAACACCAGTGCGTTTGTGGCAAT
TCCACCACCCACCCTGAGCATGCTGGACGAATACAGAGTATCTGGTCACG
ACTGCAAGAAACTGGGCTGCTAAATAAATGTGAGCGAATTCAAGGTCGAA
AAGCCAGCCTGGAGGAAATACAGCTTGTTCATTCTGAACATCACTCACTG
TTGTATGGCACCAACCCCCTGGACGGACAGAAGCTGGACCCCAGGATACT
CCTAGGTGATGACTCTCAAAAGTTTTTTTCCTCATTACCTTGTGGTGGAC
TTGGGGTGGACAGTGACACCATTTGGAATGAGCTACACTCGTCCGGTGCT
GCACGCATGGCTGTTGGCTGTGTCATCGAGCTGGCTTCCAAAGTGGCCTC
AGGAGAGCTGAAGAATGGGTTTGCTGTTGTGAGGCCCCCTGGCCATCACG
CTGAAGAATCCACAGCCATGGGGTTCTGCTTTTTTAATTCAGTTGCAATT
ACCGCCAAATACTTGAGAGACCAACTAAATATAAGCAAGATATTGATTGT
AGATCTGGATGTTCACCATGGAAACGGTACCCAGCAGGCCTTTTATGCTG
ACCCCAGCATCCTGTACATTTCACTCCATCGCTATGATGAAGGGAACTTT
TTCCCTGGCAGTGGAGCCCCAAATGAGGTTGGAACAGGCCTTGGAGAAGG
GTACAATATAAATATTGCCTGGACAGGTGGCCTTGATCCTCCCATGGGAG
ATGTTGAGTACCTTGAAGCATTCAGGAccaTCGTGAAGCCTGTGGCCAAA
GAGTTTGATCCAGACATGGTCTTAGTATCTGCTGGATTTGATGCATTGGA
AGGCCACACCCCTCCTCTAGGAGGGTACAAAGTGACGGCAAAATGTTTTG
GTCATTTGACGAAGCAATTGATGACATTGGCTGATGGACGTGTGGTGTTG
GCTCTAGAAGGAGGACATGATCTCACAGCCATCTGTGATGCATCAGAAGC
CTGTGTAAATGCCCTTCTAGGAAATGAGCTGGAGCCACTTGCAGAAGATA
TTCTCCACCAAAGCCCGAATATGAATGCTGTTATTTCTTTACAGAAGATC
ATTGAAATTCAAAGTATGTCTTTAAAGTTCTCT

FIG. 11D

(BamHI)<u>ggatcc</u>ggtaccagattacaaggacgacgatgacaagtagatcccgggtggcatccctgtgacccctccccagtg
cctctcctggccttggaagttgccactccagtgcccaccagccttgtcctaataaaattaagttgcatcattttgtctgactaggtgtc
ctctataatattatggggtggaggggggtggtatggagcaaggggcccaagttgggaagacaacctgtagggcctgcggggtc
tattcgggaaccaagctggagtgcagtggcacaatcttggctcactgcaatctccgcctcctgggttcaagcgattctcctgcctc
agcctcccgagttgttgggattccaggcatgcatgaccaggctcagctaatttttgtttttttggtagagacggggtttcaccatattg
gccaggctggtctccaactcctaatctcaggtgatctacccaccttggcctcccaaattgctgggattacaggcgtgaaccactgc
tcccttccctgtccttctgattttaaaataactataccagcaggaggacgtccagacacagcataggctacctgccatggcccaac
cggtgggacatttgagttgcttgcttggcactgtcctctcatgcgttgggtccactcagtagatgcctgttgaattgggtacgcggc
cagcttctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctca
attagtcagcaaccaggtgtggaaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagca
accatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaattttt
tttatttatgcagaggccgaggccgcctcggcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgc
aaaaagctcctcgaggaactgaaaaaccagaaagttaattccctatagtgagtcgtattaaattcgtaatcatggtcatagctgtttc
ctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagt
gagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggcc
aacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcg
gcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagca
aaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatca
caaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcg
tgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcac

FIG. 11E

gctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgc
cttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagc
agagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatct
gcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttt
ttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtg
gaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagtttta
aatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttc
gttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgata
ccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcct
gcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgt
tgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttac
atgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcact
catggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcatt
ctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaa
gtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgt
gcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagg
gaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcg
gatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgcgccctgt
agcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctccttt
cgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgc
tttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttt
gacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataa
gggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaaacgtttac
aattt

FIG. 11F

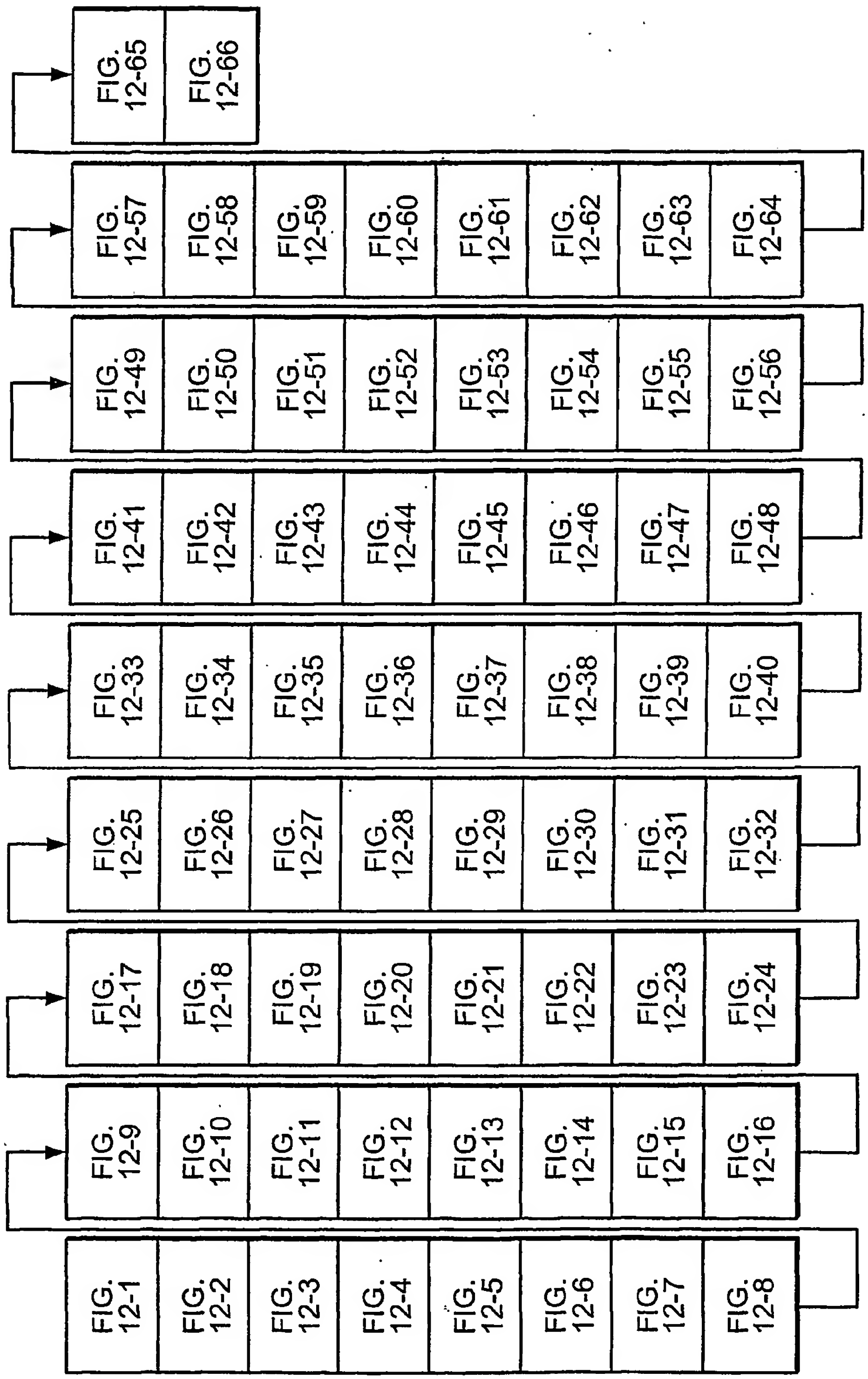


FIG. 12-1
FIG. 12-2
FIG. 12-3
FIG. 12-4
FIG. 12-5
FIG. 12-6
FIG. 12-7
FIG. 12-8
FIG. 12-9
FIG. 12-10
FIG. 12-11
FIG. 12-12
FIG. 12-13
FIG. 12-14
FIG. 12-15
FIG. 12-16
FIG. 12-17
FIG. 12-18
FIG. 12-19
FIG. 12-20
FIG. 12-21
FIG. 12-22
FIG. 12-23
FIG. 12-24
FIG. 12-25
FIG. 12-26
FIG. 12-27
FIG. 12-28
FIG. 12-29
FIG. 12-30
FIG. 12-31
FIG. 12-32
FIG. 12-33
FIG. 12-34
FIG. 12-35
FIG. 12-36
FIG. 12-37
FIG. 12-38
FIG. 12-39
FIG. 12-40
FIG. 12-41
FIG. 12-42
FIG. 12-43
FIG. 12-44
FIG. 12-45
FIG. 12-46
FIG. 12-47
FIG. 12-48
FIG. 12-49
FIG. 12-50
FIG. 12-51
FIG. 12-52
FIG. 12-53
FIG. 12-54
FIG. 12-55
FIG. 12-56
FIG. 12-57
FIG. 12-58
FIG. 12-59
FIG. 12-60
FIG. 12-61
FIG. 12-62
FIG. 12-63
FIG. 12-64
FIG. 12-65
FIG. 12-66

FIG. 12

pFLAG-CMV-5b-HDAC9

7699 base pairs

Graphic map | Table by enzyme name

```
                                          BstMCI
                  AviII                   PvuI BsiEI    EarI        MspA1I
            BglI  FspI                    BsaOI         Eam1104I     PvuII
cccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctgg
base pairs
gggtaagcggtaagtccgacgcgttgacaacccttcccgctagccacgcccggagaagcgataatgcggtcgacc
1 to 75                                                              NspBII
                  Acc16I                  BspCI         Ksp632I
                                          Bsh1285I
                                          Ple19I


cgaaagggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacg
base pairs
gctttccccctacacgacgttccgctaattcaacccattgcgggtcccaaaagggtcagtgctgcaacattttgc
76 to 150
```

FIG. 12-1

```
                                   MscI
                                 CfrI
  EaeI                        SspI MluNI
acggccagtgccaagctgatctaatcaatattggccattagccatattattcattggttatatagcataaatcaa
base pairs
tgccggtcacggttcgactagattagttataaccggtaatcggtataataagtaaccaatatatcgtatttagtt
151 to 225
  CfrI                           EaeI
                                   BalI


             MscI
             MluNI                                  SspBI
 SspI       EaeI     BsrDI                          Bsp1407I
tattggctattggccattgcatacgttgtatccatatcataatatgtacatttatattggctcatgtccaacatt
base pairs
ataaccgataaccggtaacgtatgcaacataggtatagtattatacatgtaaatataaccgagtacaggttgtaa
226 to 300
            CfrI                                    BsrGI
              BalI
```

FIG. 12-2

```
           HincII          SpeI     VspI
                                    PshBI
accgccatgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccata
base pairs
tggcggtacaactgtaactaataactgatcaataattatcattagttaatgccccagtaatcaagtatcgggtat
301 to 375
           HindII          AclNI    AsnI
                                    AseI

                                                                       Hin1I
                                                   BstMCI               AcyI
                                             BglI  BsaOI              HincII
tatggagttccgcgttacataacttacggtaaatggcccgcctggcgaccgcccagcgacccccgcccgttgacg
base pairs
atacctcaaggcgcaatgtattgaatgccatttaccgggcggaccgctggcgggtcgctgggggcgggcaactgc
376 to 450
                                                   Bsh1285I           HindII
                                                   BsiEI
```

FIG. 12-3

```
                                           BbiII
  AatII                                    Hin1I
  BbiII                                    AcyI AatII
tcaatagtgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacgg
base pairs
agttatcactgcatacaagggtatcattgcggttatccctgaaaggtaactgcagttacccacctcataaatgcc
451 to 525
  Hsp92I                                   Msp17I
                                           BsaHI
                                           Hsp92I
```

```
                                                 BbiII
                                                 Hin1I
          BglI            NdeI                   AcyI AatII
taaactgcccacttggcagtacatcaagtgtatcatatgccaagtccgccccctattgacgtcaatgacggtaaa
base pairs
atttgacgggtgaaccgtcatgtagttcacatagtatacggttcaggcgggggataactgcagttactgccattt
526 to 600
                          FauNDI                 Msp17I
                                                 BsaHI
                                                 Hsp92I
```

FIG. 12-4

```
                                                                    BstSNI
                                                                    SnaBI
tggcccgcctagcattatgcccagtacatgaccttacgggagtttcctacttggcagtacatctacgtattagtc
base pairs
accgggcggatcgtaatacgggtcatgtactggaatgccctcaaaggatgaaccgtcatgtagatgcataatcag
601 to 675
                                                                    BsaAI
                                                                   Eco105I

          NcoI Bsp19I
          StyI BstDSI
          EcoT14I
atcgctattaccatggtgatgcggttttggcagtacaccaatgggcgtggatagcggtttgactcacggggattt
base pairs
tagcgataatggtaccactacgccaaaaccgtcatgtggttacccgcacctatcgccaaactgagtgcccctaaa
676 to 750
          BssT1I
          ErhI Eco130I
          DsaI MslI
```

FIG. 12-5

```
                BbiII
                Hin1I              AccB1I
                AcyI AatII         BshNI
ccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgt
base pairs
ggttcagaggtggggtaactgcagttaccctcaaacaaaaccgtggttttagttgccctgaaaggttttacagca
751 to 825
                Msp17I             BanI
                BsaHI              Eco64I
                Hsp92I

                                                                     BanII
                                                                    Eco24I
              HincII                                               EcoICRI
aataaccccgccccgttgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctcgttta
base pairs
ttattggggcggggcaactgcgtttacccgccatccgcacatgccaccctccagatatattcgtctcgagcaaat
826 to 900
              HindII                                               Ecl136II
                                                                     SacI
```

FIG. 12-6

```
FriOI
SstI
BsiHKAI                     EagI XmaIII BstYI BspDI BcgI Eco32I
 Bbv12I     AcsI            CciNI Bsh1285I BstX2I BanIII PstI           BclI
AspHI       ApoI  HindIII BstZI BstMCI MflI Bsa29I SfcI                 Ksp22I
gtgaaccgtcagaattcaagcttgcggccgcagatctatcgatctgcaggatatcaccatgcacagtatgatcag
base pairs
cacttggcagtcttaagttcgaacgccggcgtctagatagctagacgtcctatagtggtacgtgtcatactagtc
901 to 975
 Psp124BI   EcoRI           EaeI Eco52I BglII BscI BspXI BstSFI         FbaI
                            CfrI EclXI BsiEI BseCI Bsu15I EcoRV
Alw21I                      NotI BsaOI XhoII ClaI Bsp106I

                                          FriOI            CvnI        CvnI
                                          Eco24I           AocI        AocI
                                         BpmI              Bsu36I      Bsu36I
ctcagtggatgtgaagtcagaagttcctgtgggcctggagcccatctcacctttagacctaaggacagacctcag
base pairs
gagtcacctacacttcagtcttcaaggacacccggacctcgggtagagtggaaatctggattcctgtctggagtc
976 to 1050
                                         GsuI              Eco81I      Eco81I
                                          BanII            Bse21I      Bse21I
```

FIG. 12-7

```
             DsaI         DrdI           MfeI            Asp700I
gatgatgatgcccgtggtggaccctgttgtccgtgagaagcaattgcagcaggaattacttcttatccagcagca
base pairs
ctactactacgggcaccacctgggacaacaggcactcttcgttaacgtcgtccttaatgaagaataggtcgtcgt
1051 to 1125
             BstDSI                      MunI            XmnI
```

FIG. 12-8

```
                 AlwNI
gcaacaaatccagaagcagcttctgatagcagagtttcagaaacagcatgagaacttgacacggcagcaccaggc
base pairs
cgttgtttaggtcttcgtcgaagactatcgtctcaaagtctttgtcgtactcttgaactgtgccgtcgtggtccg
1126 to 1200

BlpI
CelII     Eco57I                                                EcoNI
                                                                          AlwNI
tcagcttcaggagcatatcaaggaacttctagccataaaacagcaacaagaactcctagaaaggagcagaaact
base pairs
agtcgaagtcctcgtatagttccttgaagatcggtattttgtcgttgttcttgaggatctttcctcgtctttga
1201 to 1275
Bsp1720I
Bpu1102I
```

FIG. 12-9

```
    BpmI                                                                BseRI
ggagcagcagaggcaagaacaggaagtagagaggcatcgcagagaacagcagcttcctcctctcagaggcaaaga
base pairs
cctcgtcgtctccgttcttgtccttcatctctccgtagcgtctcttgtcgtcgaaggaggagagtctccgtttct
1276 to 1350
    GsuI                                                                 EcoNI

                                                  HindIII
tagaggacgagaaagggcagtggcaagtacagaagtaaagcag aagcttcaagagttcctactgagtaaatcagc
base pairs
atctcctgctctttcccgtcaccgttcatgtcttcatttcgtc ttcgaagttctcaaggatgactcatttagtcg
1351 to 1425
```

FIG. 12-10

```
                    Van91I                                Van91I
                    AccB7I                                AccB7I
aacgaaagacactccaactaatggaaaaatcattccgtgagccgccatcccaagctctggtacacggctgccca
base pairs
ttgctttctgtgaggttgattacctttttagtaaggcactcggcggtagggttcgagaccatgtgccgacgggt
1426 to 1500
                    Esp1396I                              Esp1396I
                    PflMI                                 PflMI


ccacacatcattggatcaaagctctccaccccttagtggaacatctccatcctacaagtacacattaccaggagc
base pairs
ggtgtgtagtaacctagtttcgagaggtggggaatcaccttgtagaggtaggatgttcatgtgtaatggtcctcg
1501 to 1575
```

FIG. 12-11

```
                          BstBI
Alw21I                    Bpu14I              FriOI
AspHI                     Csp45I              Eco24I
acaagatgcaaaggatgatttccccttcgaaaaactgcctctgagcccaacttgaaggtgcggtccaggttaaa
base pairs
tgttctacgtttcctactaaaggggaagcttttgacggagactcgggttgaacttccacgccaggtccaattt
1576 to 1650
BsiHKAI                   SfuI Bsp119I        BanII
Bbv12I                    NspV
                          LspI


                    BseRI         EcoNI
acagaaagtggcagagaggagaagcagccccttactcaggcggaaggatggaaatgttgtcacttcattcaagaa
base pairs
tgtctttcaccgtctctcctcttcgtcggggaatgagtccgccttcctacctttacaacagtgaagtaagttctt
1651 to 1725
```

FIG. 12-12

```
                                           Van91I
                                           AccB7I                Van91I
                                         BpmI PflMI              AccB7I
gcgaatgtttgaggtgacagaatcctcagtcagtagcagttctccaggctctggtcccagttcaccaaacaatgg
base pairs
cgcttacaaactccactgtcttaggagtcagtcatcgtcaagaggtccgagaccagggtcaagtggtttgttacc
1726 to 1800
                                          GsuI                 Esp1396I
                                        Esp1396I                  PflMI
                                          AlwNI

gccaactggaagtgttactgaaaatgagacttcggttttgccccctacccctcatgccgagcaaatggtttcaca
base pairs
cggttgaccttcacaatgacttttactctgaagccaaaacgggggatggggagtacggctcgtttaccaaagtgt
1801 to 1875
```

FIG. 12-13

```
          BsaMI
          Mva1269I                         BspMI                               XcmI
gcaacgcattctaattcatgaagattccatgaacctgctaagtctttatacctctccttctttgcccaacattac
base pairs
cgttgcgtaagattaagtacttctaaggtacttggacgattcagaaatatggagaggaagaaacgggttgtaatg
1876 to 1950
          BsmI RcaI
               BspHI

                                           BstBI AcsI
ErhI                                       Bpu14I
BssT1I                                     Csp45I
                                                                           Esp3I
cttggggcttcccgcagtgccatcccagctcaatgcttc gaattcactcaaagaaaagcagaagtgtgagacgca
base pairs
gaaccccgaagggcgtcacggtagggtcgagttacgaag cttaagtgagtttcttttcgtcttcacactctgcgt
1951 to 2025
EcoT14I                                    SfuI Bsp119I
                                                                           BsmBI
StyI                                       NspV ApoI
Eco130I                                    LspI EcoRI
```

FIG. 12-14

MslI
gacgcttaggcaaggtgttcctctgcctgggcagtatggaggcagcatcccggcatcttccagccaccctcatgt
base pairs
ctgcgaatccgttccacaaggagacggacccgtcatacctccgtcgtagggccgtagaaggtcggtgggagtaca
2026 to 2100

PstI
SfcI
tactttagagggaaagccacccaacagcagccaccaggctctc ctgcagcatttattattgaaagaacaaatgcg
base pairs
atgaaatctccctttcggtgggttgtcgtcggtggtccgagag gacgtcgtaaataataactttcttgtttacgc
2101 to 2175
BstSFI

FIG. 12-15

```
        HindIII                                              Eco130I
                                                              StyI
                                                              EcoT14I        ApoI
acagcaaaagcttcttgtagctggtggagttcccttacatcctcagtctcccttggcaacaaaagagagaatttc
base pairs
tgtcgttttcgaagaacatcgaccacctcaagggaatgtaggagtcagagggaaccgttgttttctctcttaaag
2176 to 2250
                                                              BssT1I         AcsI
                                                              ErhI

              Asp718I
              Acc65I
              BshNI                                                        BsgI
acctggcattagaggtacccacaaattgccccgtcacagacccctgaaccgaacccagtctgcacctttgcctca
base pairs
tggaccgtaatctccatgggtgtttaacggggcagtgtctggggacttggcttgggtcagacgtggaaacggagt
2251 to 2325
              BanI KpnI
              AccB1I
              Eco64I
```

FIG. 12-16

```
         Bpu1102I
    Alw21I Bsp1720I
    AspHI  CelII
gagcacgttggctcagctggtcattcaacagcaacaccagcaattcttggagaagcagaagcaataccagcagca
base pairs
ctcgtgcaaccgagtcgaccagtaagttgtcgttgtggtcgttaagaacctcttcgtcttcgttatggtcgtcgt
2326 to 2400
    BsiHKAI    PvuII
    Bbv12I BlpI MspA1I
               NspBII

                 BstBI
MflI             Bpu14I
XhoII            Csp45I                 Eco57I
gatccacatgaacaaactgctttcgaaatctattgaacaactgaagcaaccaggcagtcaccttgaggaagcaga
base pairs
ctaggtgtacttgtttgacgaaagctttagataacttgttgacttcgttggtccgtcagtggaactccttcgtct
2401 to 2475
BstYI            SfuI Bsp119I
BstX2I           NspV
                 LspI
```

FIG. 12-17

```
      EarI
      Eam1104I                        Bbv16II
     Asp700I                          BbsI    Bsp143II
ggaagagcttcaggggaccaggcgatgcaggaagacagagcgccctctagtggcaacagcactaggagcgacag
base pairs
ccttctcgaagtcccctggtccgctacgtccttctgtctcgcgggagatcaccgttgtcgtgatcctcgctgtc
2476 to 2550
     XmnI   Eco57I                    BpiI    HaeII
      Ksp632I                         BpuAI   BstH2I
       SapI
```

```
     BcgI
cagtgcttgtgtggatgacacactgggacaagttggggctgtgaaggtcaaggaggaaccagtggacagtgatga
base pairs
gtcacgaacacacctactgtgtgaccctgttcaaccccgacacttccagttcctccttggtcacctgtcactact
2551 to 2625
```

FIG. 12-18

```
       MflI      Van91I
       XhoII     AccB7I
agatgctcagatccaggaaatggaatctggggagcaggctgcttttatgcaacagcctttcctggaacccacgca
base pairs
tctacgagtctaggtcctttaccttagacccctcgtccgacgaaaatacgttgtcggaaaggaccttgggtgcgt
2626 to 2700
       BstYI     Esp1396I
       BstX2I    PflMI


     PmaCI
     PmlI
  AflIII                     NspBII
                                                                      Esp3I
cacacgtgcgctctctgtgcgccaagctccgctggctgcggttggcatggatggattagagaaacaccgtctcgt
base pairs
gtgtgcacgcgagagacacgcggttcgaggcgaccgacgccaaccgtacctacctaatctctttgtggcagagca
2701 to 2775
   MslI Eco72I               MspA1I
                                                                      BsmBI
     BsaAI
     BbrPI
```

FIG. 12-19

```
              EarI
     BpmI     Eam1104I                               BsrDI        BpmI
ctccaggactcactcttcccctgctgcctctgttttacctcacccagcaatggaccgcccctccagcctggctc
base pairs
gaggtcctgagtgagaaggggacgacggagacaaaatggagtgggtcgttacctggcggggaggtcggaccgag
2776 to 2850
     GsuI     Ksp632I                                             GsuI
```

```
                                                 XcmI
tgcaactggaattgcctatgaccccttgatgctgaaacaccagtgcgtttgtggcaattccaccacccaccctga
base pairs
acgttgaccttaacggatactggggaactacgactttgtggtcacgcaaacaccgttaaggtggtgggtgggact
2851 to 2925
```

FIG. 12-20

```
     SphI                                                            AcsI
     BbuI                                                            ApoI
gcatgctggacgaatacagagtatctggtcacgactgcaagaaactgggctgctaaataaatgtgagc gaattca
base pairs
cgtacgacctgcttatgtctcatagaccagtgctgacgttctttgacccgacgatttatttacactcg cttaagt
2926 to 3000
     PaeI                                                            EcoRI
     NspI

                                                                 AccB1I
                     BpmI                                         BshNI
aggtcgaaaagccagcctggaggaaatacagcttgttcattctgaacatcactcactgttgtatggcaccaaccc
base pairs
tccagcttttcggtcggacctcctttatgtcgaacaagtaagacttgtagtgagtgacaacataccgtggttggg
3001 to 3075
                     GsuI                                          BanI
                                                                  Eco64I
```

FIG. 12-21

```
                                        ErhI
                                        StyI Eco130I
BstXI              AlwNI                EcoT14I
cctggacggacagaagctggaccccaggatactcctaggtgatgactctcaaagtttttttcctcattaccttg
base pairs
ggacctgcctgtcttcgacctggggtcctatgaggatccactactgagagtttcaaaaaaaggagtaatggaac
3076 to 3150
                                        BssT1I
                                        AvrII
                                        BlnI

                                                        BsaWI        BsgI
tggtggacttggggtggacagtgacaccatttggaatgagctacactcgtccggtgctgcacgcatggctgttgg
base pairs
accacctgaaccccacctgtcactgtggtaaaccttactcgatgtgagcaggccacgacgtgcgtaccgacaacc
3151 to 3225
```

FIG. 12-22

```
                                CvnI
                                AocI                                              CfrI
                                Bsu36I          Eco57I               DraII    EaeI
ctgtgtcatcgagctggcttccaaagtggcctcaggagagctgaagaatgggtttgctgttgtgaggccccctgg
base pairs
gacacagtagctcgaccgaaggtttcaccggagtcctctcgacttcttacccaaacgacaacactccgggggacc
3226 to 3300
                                Eco81I                                    EcoO109I
                                Bse21I

MscI                      ErhI Eco130I
                          BssT1I  BstXI
                Eco57I   MslI DsaI
ccatcacgctgaagaatccacagccatggggttctgctttttaattcagttgcaattaccgccaaatacttgag
base pairs
ggtagtgcgacttcttaggtgtcggtaccccaagacgaaaaattaagtcaacgttaatggcggtttatgaactc
3301 to 3375
MluNI                     EcoT14I
BalI                      StyI BstDSI
                          NcoI Bsp19I
```

FIG. 12-23

```
                        BstX2I        NcoI Bsp19I Asp718I    SseBI
                         BstYI        StyI BstDSI AccB1I
                                                       Eco147I
   BsaI                  XhoII        EcoT14I  BshNI        StuI
agaccaactaaatataagcaagatattgattgtagatctggatgttcaccatggaaacggtacccagcaggcctt
base pairs
tctggttgatttatattcgttctataactaacatctagacctacaagtggtacctttgccatgggtcgtccggaa
3376 to 3450
   Eco31I                BglII        BssT1I    BanI KpnI   AatI
                         MflI         ErhI Eco130I Eco64I   Pme55I
                                      DsaI      Acc65I
```

```
            SspBI
            Bsp1407I          MslI              Asp700I
ttatgctgaccccagcatcctgtacatttcactccatcgctatgatgaagggaactttttccctggcagtggagc
base pairs
aatacgactggggtcgtaggacatgtaaagtgaggtagcgatactacttcccttgaaaaagggaccgtcacctcg
3451 to 3525
            BsrGI                               XmnI
```

FIG. 12-24

```
                    SseBI ErhI
FriOI               Eco147I
Eco24I              StuI BssT1I                 SspI
cccaaatgaggttggaacaggccttggagaagggtacaatataaatattgcctggacaggtggccttgatcctcc
base pairs
gggtttactccaaccttgtccggaacctcttcccatgttatatttataacggacctgtccaccggaactaggagg
3526 to 3600
BanII               AatI StyI
                    Pme55I Eco130I
                     EcoT14I

NcoI Bsp19I                                        MscI
StyI BstDSI                   BsaMI                MluNI
                                                                  AspI
EcoT14I                       Mva1269I            EaeI
                                                                  AtsI
catgggagatgttgagtaccttgaagcattcaggaccatcgtgaagcctgtggccaaagagtttgatccagacat
base pairs
gtaccctctacaactcatggaacttcgtaagtcctggtagcacttcggacaccggtttctcaaactaggtctgta
3601 to 3675
BssT1I                        BsmI                 CfrI
DsaI                                                            Tth111I
ErhI Eco130I                                        BalI
```

FIG. 12-25

```
              Mph1103I
              EcoT22I
           Ppu10I                 EcoNI
ggtcttagtatctgctggatttgatgcattggaaggccacacccctcctctaggagggtacaaagtgacggcaaa
base pairs
ccagaatcatagacgacctaaactacgtaaccttccggtgtggggaggagatcctcccatgtttcactgccgttt
3676 to 3750
              NsiI                 BseRI
              Zsp2I

           MfeI                  AflIII        XbaI
atgttttggtcatttgacgaagcaattgatgacattggctgatggacgtgtggtgttggctctagaaggaggaca
base pairs
tacaaaaccagtaaactgcttcgttaactactgtaaccgactacctgcacaccacaaccgagatcttcctcctgt
3751 to 3825
           MunI
```

FIG. 12-26

```
             Mph1103I
             EcoT22I
          Ppu10I                                           BpmI
tgatctcacagccatctgtgatgcatcagaagcctgtgtaaatgcccttctaggaaatgagctggagccacttgc
base pairs
actagagtgtcggtagacactacgtagtcttcggacacatttacgggaagatcctttactcgacctcggtgaacg
3826 to 3900
              NsiI                                         GsuI
              Zsp2I

                                   BsaMI
     Asp700I                       Mva1269I                  ApoI
agaagatattctccaccaaagcccgaatatgaatgctgttatttctttacagaagatcattgaaattcaaagtat
base pairs
tcttctataagaggtggtttcgggcttatacttacgacaataaagaaatgtcttctagtaactttaagtttcata
3901 to 3975
     XmnI                          BsmI                      AcsI
```

FIG. 12-27

```
                 MflI   AccB1I                                  AvaI  BcoI
                 BstI  BsaWI  KpnI                          MflI  Eco88I  PspALI
      DraI       BamHI  BshNI                               XhoII  Cfr9I  SmaI  MslI
gtctttaaagttctctggatccggtaccagattacaaggacgacgatgacaagtagatcccgggtggcatccctg
base pairs
cagaaatttcaagagacctaggccatggtctaatgttcctgctgctactgttcatctagggcccaccgtagggac
3976 to 4050
                 XhoII  BanI  Eco64I                          BstYI  Ama87I
                 BstYI  Acc65I                                BstX2I  BsoBI
                 BstX2I  Asp718I                                  XmaI  PspAI

                          Eco130I
                          StyI                 GsuI
                          EcoT14I             MslI
tgacccctccccagtgcctctcctggccttggaagttgccactccagtgcccaccagccttgtcctaataaaatt
base pairs
actggggaggggtcacggagaggaccggaaccttcaacggtgaggtcacgggtggtcggaacaggattattttaa
4051 to 4125
                          BssT1I               BpmI
                          ErhI
```

FIG. 12-28

```
                          AspEI                                    DraII
                         Eam1105I                     SspI        PspOMI
aagttgcatcattttgtctgactaggtgtcctctataatattatggggtggagggggtggtatggagcaagggg
base pairs
ttcaacgtagtaaaacagactgatccacaggagatattataataccccacctcccccaccatacctcgttcccc
4126 to 4200
                          EclHKI
                                                                   Bsp120I
                          AhdI
                                                                     EcoO1

  Eco24I             SfcI
    BanII         Bbv16II
     FriOI        BbsI     DraII                             BpmI BsgI
cccaagttgggaagacaacctgtagggcctgcggggtctattcgggaaccaagctggagtgcagtggcacaatct
base pairs
gggttcaacccttctgttggacatcccggacgccccagataagcccttggttcgacctcacgtcaccgtgttaga
4201 to 4275
                  BpiI     EcoO109I                          GsuI
   09I            BpuAI
  ApaI                BstSFI
```

FIG. 12-29

```
                                                BcoI
                                               Ama87I
                                        BcgI    AvaI
tggctcactgcaatctccgcctcctgggttcaagcgattctcctgcctcagcctcccgagttgttgggattccag
base pairs
accgagtgacgttagaggcggaggacccaagttcgctaagaggacggagtcggagggctcaacaaccctaaggtc
4276 to 4350
                                               Eco88I
                                               BsoBI

     NspI       BlpI                                                 MscI
    PaeI Mph1103I                                                   MluNI
  Ppu10I EcoT22I                         Esp3I                      EaeI
gcatgcatgaccaggctcagctaatttttgtttttttggtagagacggggtttcaccatattggccaggctggtc
base pairs
cgtacgtactggtccgagtcgattaaaaacaaaaaaaccatctctgccccaaagtggtataaccggtccgaccag
4351 to 4425
    BbuI Zsp2I CelII                      BsmBI               CfrI
    SphI       Bsp1720I                                        BalI
      NsiI     Bpu1102I
```

FIG. 12-30

```
                                   Eco130I
                                    StyI
  BsaI                             EcoT14I          BstXI
tccaactcctaatctcaggtgatctacccaccttggcctcccaaattgctgggattacaggcgtgaaccactgct
base pairs
aggttgaggattagagtccactagatgggtggaaccggagggtttaacgaccctaatgtccgcacttggtgacga
4426 to 4500
  Eco31I                           BssT1I
                                    ErhI
```

FIG. 12-31

```
                                        BbiII
                                        Hin1I                NcoI
                                                             StyI
            DraI                        AcyI  AatII          EcoT14I
cccttccctgtccttctgattttaaaataactataccagcaggaggacgtccagacacagcataggctacctgcc
base pairs
gggaagggacaggaagactaaaattttattgatatggtcgtcctcctgcaggtctgtgtcgtatccgatggacgg
4501 to 4575
                                        Msp17I               BssT1I

                                        BsaHI                ErhI

                                                             BspMI
                                        Hsp92I


   Eco130I BsrFI PflMI
   DsaI AgeI Bse118I
       BsaWI AccB7I
atggcccaaccggtgggacatttgagttgcttgcttggcactgtcctctcatgcgttgggtccactcagtagatg
base pairs
taccgggttggccaccctgtaaactcaacgaacgaaccgtgacaggagagtacgcaacccaggtgagtcatctac
4576 to 4650
       BssAI Esp1396I
   BstDSI PinAI Van91I
    Bsp19I Cfr10I
```

FIG. 12-32

```
                    EaeI      AlwNI
cctgttgaattgggtacgcggccagcttctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctcccc
base pairs
ggacaacttaacccatgcgccggtcgaagacaccttacacacagtcaatcccacacctttcaggggtccgagggg
4651 to 4725
                    CfrI


                          NspI
                          PaeI Mph1103I
                        Ppu10I EcoT22I       SexAI
agcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaaagtccccaggctccccag
base pairs
tcgtccgtcttcatacgtttcgtacgtagagttaatcagtcgttggtccacaccttttcaggggtccgaggggtc
4726 to 4800
                          BbuI Zsp2I
                          SphI
                            NsiI
```

FIG. 12-33

```
                    NspI
                     PaeI Mph1103I
                   Ppu10I EcoT22I
caggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgc
base pairs
gtccgtcttcatacgtttcgtacgtagagttaatcagtcgttggtatcagggcggggattgaggcgggtagggcg
4801 to 4875
                    BbuI Zsp2I
                    SphI
                      NsiI

                              NcoI Bsp19I
                              StyI  BstDSI
                              EcoT14I
ccctaactccgcccagttccgcccattctccgccccatggctgactaattttttttatttatgcagaggccgagg
base pairs
gggattgaggcgggtcaaggcgggtaagaggcggggtaccgactgattaaaaaaaataaatacgtctccggctcc
4876 to 4950
                              BssT1I
                              ErhI Eco130I
                              DsaI
```

FIG. 12-34

```
                                                SseBI AvrII
                                               Eco147I  BlnI
     BglI                              BseRI    StuI BssT1I
ccgcctcggcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgcaaaaagctc c
base pairs
ggcggagccggagactcgataaggtcttcatcactcctccgaaaaacctccggatccgaaaacgtttttcgagg
4951 to 5025
     SfiI                                       AatI StyI
                                               Pme55I  ErhI
                                                EcoT14I Eco130I

Ama87I
Eco88I BseRI
AvaI BsoBI                       SfcI           ApoI
tcgaggaactgaaaaaccagaaagttaattccctatagtgagtcgtattaaattcgtaatcatggtcatagctgt
base pairs
agctccttgactttttggtctttcaattaagggatatcactcagcataatttaagcattagtaccagtatcgaca
5026 to 5100
XhoI BcoI                        BstSFI         AcsI
Sfr274I
PaeR7I
```

FIG. 12-35

```
                           AccBSI
                           BsrBI
ttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggg
base pairs
aaggacacactttaacaataggcgagtgttaaggtgtgttgtatgctcggccttcgtatttcacatttcggaccc
5101 to 5175
                           BstD102I

AccB1I                       VspI
BshNI                        PshBI
gtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgt
base pairs
cacggattactcactcgattgagtgtaattaacgcaacgcgagtgacgggcgaaaggtcagccctttggacagca
5176 to 5250
BanI                         AsnI
Eco64I                       AseI
```

FIG. 12-36

```
        VspI                                                     Eam1104I
   MspA1I                                                    BstH2I
   PvuII PshBI    EaeI                                       Bsp143II
gccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgc
base pairs
cggtcgacgtaattacttagccggttgcgcgcccctctccgccaaacgcataacccgcgagaaggcgaaggagcg
5251 to 5325
   NspBII         CfrI                                       HaeII EarI
        AsnI                                                     SapI
        AseI                                                    Ksp632I

                BstMCI          AccBSI
                BsaOI           BsrBI
tcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttat
base pairs
agtgactgagcgacgcgagccagcaagccgacgccgctcgccatagtcgagtgagtttccgccattatgccaata
5326 to 5400
                Bsh1285I        BstD102I
                BsiEI
```

FIG. 12-37

NspI
BspLU11I
ccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaagg
base pairs
ggtgtcttagtcccctattgcgtcctttcttgtacactcgttttccggtcgttttccggtccttggcatttttcc
5401 to 5475
AflIII

DrdI
ccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggt
base pairs
ggcgcaacgaccgcaaaaaggtatccgaggcggggggactgctcgtagtgtttttagctgcgagttcagtctcca
5476 to 5550

FIG. 12-38

```
                                                                      BsiI
ggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccga
base pairs
ccgctttgggctgtcctgatatttctatggtccgcaaagggggaccttcgagggagcacgcgagaggacaaggct
5551 to 5625
                                                                      BssSI


                                                                BstH2I
            BsaWI                                               Bsp143II          SfcI
ccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgta
base pairs
gggacggcgaatggcctatggacaggcggaaagagggaagcccttcgcaccgcgaaagagttacgagtgcgacat
5626 to 5700
                                                                HaeII             BstSFI
```

FIG. 12-39

```
                                          BsiHKAI
                                                        NspBII
                                      Alw44I            BstMCI
                                      VneI Bbv12I        BsaOI
ggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgct
base pairs
ccatagagtcaagccacatccagcaagcgaggttcgacccgacacacgtgcttggggggcaagtcgggctggcga
5701 to 5775
                                      ApaLI
                                                       Bsh1285I
                                         AspHI          BsiEI
                                         Alw21I
                                                        MspA1I

      BsaWI                                                     AlwNI
gcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactg
base pairs
cgcggaataggccattgatagcagaactcaggttgggccattctgtgctgaatagcggtgaccgtcgtcggtgac
5776 to 5850
```

FIG. 12-40

SfcI

gtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctaca
base pairs
cattgtcctaatcgtctcgctccatacatccgccacgatgtctcaagaacttcaccaccggattgatgccgatgt
5851 to 5925

BstSFI

Eco57I

ctagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgat
base pairs
gatcttcttgtcataaaccatagacgcgagacgacttcggtcaatggaagcctttttctcaaccatcgagaacta
5926 to 6000

FIG. 12-41

```
                                                                MflI
              NspBII                                            XhoII
ccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggat
base pairs
ggccgtttgtttggtggcgaccatcgccaccaaaaaaacaaacgttcgtcgtctaatgcgcgtctttttttccta
6001 to 6075
              MspA1I                                            BstYI
                                                               BstX2I

      MflI
      XhoII
ctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttgg
base pairs
gagttcttctaggaaactagaaaagatgccccagactgcgagtcaccttgcttttgagtgcaattccctaaaacc
6076 to 6150
      BstYI
      BstX2I
```

FIG. 12-42

```
                 MflI         MflI
 RcaI            XhoII        XhoII     DraI              DraI
tcatgagattatcaaaaaggatcttcacctagatcctttaaattaaaaatgaagttttaaatcaatctaaagta
base pairs
agtactctaatagtttttcctagaagtggatctaggaaatttaatttttacttcaaaatttagttagatttcat
6151 to 6225
 BspHI           BstYI        BstYI
                 BstX2I       BstX2I
```

```
                                            AccB1I
                                            BshNI
tatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttc
base pairs
atatactcatttgaaccagactgtcaatggttacgaattagtcactccgtggatagagtcgctagacagataaag
6226 to 6300
                                            BanI
                                            Eco64I
```

FIG. 12-43

```
                  Eam1105I
                  AspEI
gttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtg
base pairs
caagtaggtatcaacggactgaggggcagcacatctattgatgctatgccctcccgaatggtagaccggggtcac
6301 to 6375
                  EclHKI
                  AhdI

                                    Cfr10I
        BsrDI            BsaI       BssAI      BpmI                          BglI
ctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccg
base pairs
gacgttactatggcgctctgggtgcgagtggccgaggtctaaatagtcgttatttggtcggtcggccttcccggc
6376 to 6450
                         Eco31I     BsrFI      GsuI
                                    Bse118I
```

FIG. 12-44

```
                                                  VspI
                                                  PshBI
agcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagta
base pairs
tcgcgtcttcaccaggacgttgaaataggcggaggtaggtcagataattaacaacggcccttcgatctcattcat
6451 to 6525
                                                  AsnI
                                                  AseI

                   AviII                BstSFI
                   FspI                 SfcI      MslI
gttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggta
base pairs
caagcggtcaattatcaaacgcgttgcaacaacggtaacgatgtccgtagcaccacagtgcgagcagcaaaccat
6526 to 6600
                   Acc16I               BsrDI
                       Psp1406I
```

FIG. 12-45

```
              BsaWI
tggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggtta
base pairs
accgaagtaagtcgaggccaagggttgctagttccgctcaatgtactagggggtacaacacgttttttcgccaat
6601 to 6675

                 BstMCI
                 PvuI BsiEI
                 BsaOI              EaeI                      MslI
gctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgc
base pairs
cgaggaagccaggaggctagcaacagtcttcattcaaccggcgtcacaatagtgagtaccaataccgtcgtgacg
6676 to 6750
                 BspCI              CfrI
                 Bsh1285I
                 Ple19I
```

FIG. 12-46

Acc113I
Eco255I
ataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgag
base pairs
tattaagagaatgacagtacggtaggcattctacgaaaagacactgaccactcatgagttggttcagtaagactc
6751 to 6825
ScaI

BbiII
BstMCI HinlI
BsaOI BcgI AcyI
aatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactt
base pairs
ttatcacatacgccgctggctcaacgagaacgggccgcagttatgccctattatggcgcggtgtatcgtcttgaa
6826 to 6900
Bsh1285I Msp17I
BsiEI BsaHI
Hsp92I

FIG. 12-47

```
        Alw21I      XmnI                        MflI              MflI
 DraI   AspHI       Psp1406I                    XhoII     NspBII  XhoII
taaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagtt
base pairs
attttcacgagtagtaaccttttgcaagaagccccgcttttgagagttcctagaatggcgacaactctaggtcaa
6901 to 6975
        BsiHKAI     Asp700I                     BstYI     MspA1I  BstYI
        Bbv12I                                  BstX2I            BstX2I

            BssSI
           Alw44I Bbv12I
           VneI BsiHKAI         Eco57I
cgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaa
base pairs
gctacattgggtgagcacgtgggttgactagaagtcgtagaaaatgaaagtggtcgcaaagacccactcgttttt
6976 to 7050
           ApaLI Alw21I
             BsiI
               AspHI
```

FIG. 12-48

```
                                                                 EarI
                                          MslI                   Eam1104I
caggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttc
base pairs
gtccttccgttttacggcgtttttcccttattcccgctgtgcctttacaacttatgagtatgagaaggaaaaag
7051 to 7125
                                                                 Ksp632I

                                                 AccBSI
   SspI                                RcaI      BsrBI
aatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaac
base pairs
ttataataacttcgtaaatagtcccaataacagagtactcgcctatgtataaacttacataaatctttttatttg
7126 to 7200
                                       BspHI     BstD102I
```

FIG. 12-49

```
                                                          SfcI
aaataggggttccgcgcacatttccccgaaaagtgccacctgacgcgccctgtagcggcgcattaagcgcggcgg
base pairs
tttatccccaaggcgcgtgtaaaggggcttttcacggtggactgcgcgggacatcgccgcgtaattcgcgccgcc
7201 to 7275
                                                          BstSFI

                                                                 AccBSI
                                                 BstH2I   HaeII BstD102I
                                                 Bsp143II       BsrBI
gtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttccctt
base pairs
cacaccaccaatgcgcgtcgcactggcgatgtgaacggtcgcgggatcgcgggcgaggaaagcgaaagaaggga a
7276 to 7350
                                                  HaeII   Bsp143II
                                                          BstH2I
```

FIG. 12-50

```
                BsrFI
                BssAI NaeI
                MroNI Bse118I
cctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtg
base pairs
ggaaagagcggtgcaagcggccgaaaggggcagttcgagatttagccccgtagggaaatcccaaggctaaatcac
7351 to 7425
                NgoAIV
                NgoMI
                Cfr10I


       AccB1I
       BshNI                                    BsaAI
ctttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacgg
base pairs
gaaatgccgtggagctggggtttttgaactaatcccactaccaagtgcatcacccggtagcgggactatctgcc
7426 to 7500
       BanI                                        DraIII
       Eco64I
```

FIG. 12-51

```
                    DrdI
tttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaacc
base pairs
aaaaagcgggaaactgcaacctcaggtgcaagaaattatcacctgagaacaaggtttgaccttgttgtgagttgg
7501 to 7575
```

```
ctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgattt
base pairs
gatagagccagataagaaaactaaatattccctaaaacggctaaagccggataaccaattttttactcgactaaa
7576 to 7650
```

FIG. 12-52

```
     ApoI       ApoI           SspI  Psp1406I
aacaaaaatttaacgcgaattttaacaaaatattaaacgtttacaattt  base pairs
ttgtttttaaattgcgcttaaaattgttttataatttgcaaatgttaaa  7651 to 7699
     AcsI       AcsI
```

Table by Enzyme Name

Enzyme name	No. cuts	Positions of sites	Recognition sequence	
AatI	3	3446 3546 5002	agg/cct	More info
AatII	5	451 504 587 773 4550	gacgt/c	More info
Acc113I	1	6804	agt/act	More info
Acc16I	2	21 6546	tgc/gca	More info
Acc65I	3	2264 3434 3998	g/ gtacc	More info
AccB1I	8	791 2264 3065 3434 3998 5175 6272 7432	g/ gyrcc	More info
AccB7I	6	1445 1482 1775 1796 2644 4587	ccannnn/ntgg	More info
AccBSI	4	5126 5367 7168 7332	gagcgg	More info
AclNI	1	326	a/ ctagt	More info
AcsI	8	912 1990 2244 2994 3963 5075 7656 7667	r/ aatty	More info
AcyI	6	448 501 584 770 4547 6861	gr/cgyc	More info

FIG. 12-53

AflIII	3	2702 3796 5431	a/ crygt	More info
AgeI	1	4584	a/ ccggt	More info
AhdI	2	4150 6324	gacnnn/nngtc	More info
Alw21I	6	894 1576 2330 5749 6910 6995	gwgcw/c	More info
Alw44I	2	5745 6991	g/ tgcac	More info
AlwNI	6	1147 1273 1775 3091 4678 5847	cagnnn/ctg	More info
Ama87I	3	4034 4330 5025	c/ ycgrg	More info
AocI	3	1034 1046 3256	cc/ tnagg	More info
ApaI	1	4202	gggcc/c	More info
ApaLI	2	5745 6991	g/ tgcac	More info
ApoI	8	912 1990 2244 2994 3963 5075 7656 7667	r/ aatty	More info
AseI	4	334 5202 5261 6496	at/ taat	More info
AsnI	4	334 5202 5261 6496	at/ taat	More info
Asp700I	5	1107 2481 3506 3906 6923	gaann/nnttc	More info
Asp718I	3	2264 3434 3998	g/ gtacc	More info

FIG. 12-54

AspEI	2	4150 6324	gacnnn/nngtc	More info
AspHI	6	894 1576 2330 5749 6910 6995	gwgcw/c	More info
AspI	1	3674	gacn/nngtc	More info
AtsI	1	3674	gacn/nngtc	More info
AvaI	3	4034 4330 5025	c/ ycgrg	More info
AviII	2	21 6546	tgc/gca	More info
AvrII	2	3109 5003	c/ ctagg	More info
BalI	5	184 238 3300 3653 4414	tgg/cca	More info
BamHI	1	3992	g/ gatcc	More info
BanI	8	791 2264 3065 3434 3998 5175 6272 7432	g/ gyrcc	More info
BanII	5	894 1017 1623 3526 4202	grgcy/c	More info
BanIII	1	939	at/ cgat	More info
BbiII	6	448 501 584 770 4547 6861	gr/cgyc	More info
BbrPI	1	2705	cac/gtg	More info
BbsI	2	2512 4216	gaagac	More info
BbuI	4	2930 4355 4750 4823	gcatg/c	More info
Bbv12I	6	894 1576 2330 5749 6910 6995	gwgcw/c	More info
Bbv16II	2	2512 4216	gaagac	More info
BcgI	4	941 2556 4321 6851	cgannnnnntgc	More info

FIG. 12-55

BclI	1	969	t/ gatca	More info
BcoI	3	4034 4330 5025	c/ ycgrg	More info
BglI	5	14 417 538 4956 6444	gccnnnn/nggc	More info
BglII	2	932 3409	a/ gatct	More info
BlnI	2	3109 5003	c/ ctagg	More info
BlpI	3	1200 2337 4366	gc/tnagc	More info
BpiI	2	2512 4216	gaagac	More info
BpmI	10	1015 1279 1772 2781 2842 3022 3892 4097 4259 6414	ctggag	More info
Bpu1102I	3	1200 2337 4366	gc/tnagc	More info
Bpu14I	3	1603 1988 2423	tt/cgaa	More info
BpuAI	2	2512 4216	gaagac	More info
Bsa29I	1	939	at/ cgat	More info
BsaAI	3	666 2705 7473	yac/gtr	More info
BsaHI	6	448 501 584 770 4547 6861	gr/cgyc	More info
BsaI	3	3380 4427 6396	ggtctc	More info
BsaMI	3	1886 3631 3936	gaatgc	More info
BsaOI	7	42 424 928 5347 5771 6694 6843	cgry/cg	More info
BsaWI	6	3200 3995 4584 5637 5784 6615	w/ ccggw	More info
BscI	1	939	at/ cgat	More info

FIG. 12-56

Bse118I	3	4584 6404 7368	r/ ccggy	More info
Bse21I	3	1034 1046 3256	cc/ tnagg	More info
BseCI	1	939	at/ cgat	More info
BseRI	5	1337 1671 3725 4989 5027	gaggag	More info
BsgI	3	2315 3212 4264	gtgcag	More info
Bsh1285I	7	42 424 928 5347 5771 6694 6843	cgry/cg	More info
BshNI	8	791 2264 3065 3434 3998 5175 6272 7432	g/ gyrcc	More info
BsiEI	7	42 424 928 5347 5771 6694 6843	cgry/cg	More info
BsiHKAI	6	894 1576 2330 5749 6910 6995	gwgcw/c	More info
BsiI	2	5609 6993	ctcgtg	More info
BsmBI	3	2023 2773 4397	cgtctc	More info
BsmI	3	1886 3631 3936	gaatgc	More info
BsoBI	3	4034 4330 5025	c/ ycgrg	More info
Bsp106I	1	939	at/ cgat	More info
Bsp119I	3	1603 1988 2423	tt/cgaa	More info
Bsp120I	1	4198	g/ ggccc	More info
Bsp1407I	2	270 3471	t/ gtaca	More info
Bsp143II	5	2519 5309 5679 7318 7326	rgcgc/y	More info
Bsp1720I	3	1200 2337 4366	gc/tnagc	More info
Bsp19I	6	686 3324 3424 3600 4574 4910	c/ catgg	More info

FIG. 12-57

BspCI	2	42 6694	cgat/cg	More info
BspDI	1	939	at/ cgat	More info
BspHI	3	1891 6151 7159	t/ catga	More info
BspLU11I	1	5431	a/ catgt	More info
BspMI	2	1913 4574	acctgc	More info
BspXI	1	939	at/ cgat	More info
BsrBI	4	5126 5367 7168 7332	gagcgg	More info
BsrDI	4	245 2827 6383 6565	gcaatg	More info
BsrFI	3	4584 6404 7368	r/ ccggy	More info
BsrGI	2	270 3471	t/ gtaca	More info
BssAI	3	4584 6404 7368	r/ ccggy	More info
BssSI	2	5609 6993	ctcgtg	More info
BssT1I	13	686 1950 2226 3109 3324 3424 3547 3600 4077 4456 4574 4910 5003	c/ cwwgg	More info
BstBI	3	1603 1988 2423	tt/cgaa	More info
BstD102I	4	5126 5367 7168 7332	gagcgg	More info
BstDSI	7	686 1062 3324 3424 3600 4574 4910	c/ crygg	More info
BstH2I	5	2519 5309 5679 7318 7326	rgcgc/y	More info

FIG. 12-58

BstI	1	3992	g/ gatcc	More info
BstMCI	7	42 424 928 5347 5771 6694 6843	cgry/cg	More info
BstSFI	8	944 2144 4220 5058 5696 5887 6565 7250	c/ tryag	More info
BstSNI	1	666	tac/gta	More info
BstX2I	12	932 2400 2634 3409 3992 4030 6072 6083 6169 6181 6949 6966	r/ gatcy	More info
BstXI	3	3076 3325 4473	ccannnnn/ntgg	More info
BstYI	12	932 2400 2634 3409 3992 4030 6072 6083 6169 6181 6949 6966	r/ gatcy	More info
BstZI	1	925	c/ ggccg	More info
Bsu15I	1	939	at/ cgat	More info
Bsu36I	3	1034 1046 3256	cc/ tnagg	More info
CciNI	1	925	gc/ggccgc	More info
CelII	3	1200 2337 4366	gc/tnagc	More info
Cfr10I	3	4584 6404 7368	r/ ccggy	More info
Cfr9I	1	4034	c/ ccggg	More info
CfrI	10	152 182 236 925 3298 3651 4412 4669 5270 6712	y/ ggccr	More info
ClaI	1	939	at/ cgat	More info
Csp45I	3	1603 1988 2423	tt/cgaa	More info
CvnI	3	1034 1046 3256	cc/ tnagg	More info

FIG. 12-59

DraI	5	3981 4523 6190 6209 6901	ttt/aaa	More info
DraII	3	3291 4198 4225	rg/gnccy	More info
DraIII	1	7476	cacnnn/gtg	More info
DrdI	3	1076 5539 7520	gacnnnn/nngtc	More info
DsaI	7	686 1062 3324 3424 3600 4574 4910	c/ crygg	More info
EaeI	10	152 182 236 925 3298 3651 4412 4669 5270 6712	y/ ggccr	More info
EagI	1	925	c/ ggccg	More info
Eam1104I	5	58 2482 2793 5314 7118	ctcttc	More info
Eam1105I	2	4150 6324	gacnnn/nngtc	More info
EarI	5	58 2482 2793 5314 7118	ctcttc	More info
Ecl136II	1	892	gag/ ctc	More info
EclHKI	2	4150 6324	gacnnn/nngtc	More info
EclXI	1	925	c/ ggccg	More info
Eco105I	1	666	tac/gta	More info
Eco130I	13	686 1950 2226 3109 3324 3424 3547 3600 4077 4456 4574 4910 5003	c/ cwwgg	More info
Eco147I	3	3446 3546 5002	agg/cct	More info

FIG. 12-60

Eco24I	5	894 1017 1623 3526 4202	grgcy/c	More info
Eco255I	1	6804	agt/act	More info
Eco31I	3	3380 4427 6396	ggtctc	More info
Eco32I	1	952	gat/ atc	More info
Eco52I	1	925	c/ ggccg	More info
Eco57I	7	1210 2446 2488 3271 3314 5963 7011	ctgaag	More info
Eco64I	8	791 2264 3065 3434 3998 5175 6272 7432	g/ gyrcc	More info
Eco72I	1	2705	cac/gtg	More info
Eco81I	3	1034 1046 3256	cc/ tnagg	More info
Eco88I	3	4034 4330 5025	c/ ycgrg	More info
EcoICRI	1	892	gag/ ctc	More info
EcoNI	4	1259 1338 1684 3723	cctnn/nnnagg	More info
EcoO109I	3	3291 4198 4225	rg/gnccy	More info
EcoRI	3	912 1990 2994	g/ aattc	More info
EcoRV	1	952	gat/ atc	More info
EcoT14I	13	686 1950 2226 3109 3324 3424 3547 3600 4077 4456 4574 4910 5003	c/ cwwgg	More info
EcoT22I	5	3703 3850 4357 4752 4825	atgca/t	More info

FIG. 12-61

ErhI	13	686 1950 2226 3109 3324 3424 3547 3600 4077 4456 4574 4910 5003	c/ cwwgg	More info
Esp1396I	6	1445 1482 1775 1796 2644 4587	ccannnn/ntgg	More info
Esp3I	3	2023 2773 4397	cgtctc	More info
FauNDI	1	560	ca/ tatg	More info
FbaI	1	969	t/ gatca	More info
FriOI	5	894 1017 1623 3526 4202	grgcy/c	More info
FspI	2	21 6546	tgc/gca	More info
GsuI	10	1015 1279 1772 2781 2842 3022 3892 4097 4259 6414	ctggag	More info
HaeII	5	2519 5309 5679 7318 7326	rgcgc/y	More info
Hin1I	6	448 501 584 770 4547 6861	gr/cgyc	More info
HincII	3	311 446 842	gty/rac	More info
HindII	3	311 446 842	gty/rac	More info
HindIII	3	918 1394 2183	a/ agctt	More info
Hsp92I	6	448 501 584 770 4547 6861	gr/cgyc	More info
KpnI	3	2268 3438 4002	ggtac/c	More info
Ksp22I	1	969	t/ gatca	More info
Ksp632I	5	58 2482 2793 5314 7118	ctcttc	More info
LspI	3	1603 1988 2423	tt/cgaa	More info
MfeI	2	1091 3773	c/ aattg	More info
MflI	12	932 2400 2634 3409 3992 4030	r/gatcy	More info

FIG. 12-62

		6072 6083 6169 6181 6949 6966		
MluNI	5	184 238 3300 3653 4414	tgg/cca	More info
Mph1103I	5	3703 3850 4357 4752 4825	atgca/t	More info
MroNI	1	7368	g/ ccggc	More info
MscI	5	184 238 3300 3653 4414	tgg/cca	More info
MslI	10	691 2094 2703 3323 3489 4047 4094 6576 6735 7094	caynn/nnrtg	More info
Msp17I	6	448 501 584 770 4547 6861	gr/cgyc	More info
MspA1I	7	71 2341 2731 5255 5773 6018 6959	cmg/ckg	More info
MunI	2	1091 3773	c/ aattg	More info
Mva1269I	3	1886 3631 3936	gaatgc	More info
NaeI	1	7370	gcc/ggc	More info
NcoI	6	686 3324 3424 3600 4574 4910	c/ catgg	More info
NdeI	1	560	ca/ tatg	More info
NgoAIV	1	7368	g/ ccggc	More info
NgoMI	1	7368	g/ ccggc	More info
NotI	1	925	gc/ggccgc	More info
NsiI	5	3703 3850 4357 4752 4825	atgca/t	More info
NspBII	7	71 2341 2731 5255 5773 6018 6959	cmg/ckg	More info
NspI	5	2930 4355 4750 4823 5435	rcatg/y	More info
NspV	3	1603 1988 2423	tt/cgaa	More info
PaeI	4	2930 4355 4750 4823	gcatg/c	More info
PaeR7I	1	5025	c/ tcgag	More info

FIG. 12-63

PflMI	6	1445 1482 1775 1796 2644 4587	ccannnn/ntgg	More info
PinAI	1	4584	a/ ccggt	More info
Ple19I	2	42 6694	cgat/cg	More info
PmaCI	1	2705	cac/gtg	More info
Pme55I	3	3446 3546 5002	agg/cct	More info
PmlI	1	2705	cac/gtg	More info
Ppu10I	5	3699 3846 4353 4748 4821	a/ tgcat	More info
PshBI	4	334 5202 5261 6496	at/ taat	More info
Psp124BI	1	894	gagct/c	More info
Psp1406I	3	6550 6923 7687	aa/cgtt	More info
PspAI	1	4034	c/ ccggg	More info
PspALI	1	4036	ccc/ggg	More info
PspOMI	1	4198	g/ ggccc	More info
PstI	2	948 2148	ctgca/g	More info
PvuI	2	42 6694	cgat/cg	More info
PvuII	3	71 2341 5255	cag/ctg	More info
RcaI	3	1891 6151 7159	t/ catga	More info
SacI	1	894	gagct/c	More info
SapI	2	2483 5314	gctcttc	More info
ScaI	1	6804	agt/act	More info
SexAI	1	4769	a/ ccwggt	More info
SfcI	8	944 2144 4220 5058 5696 5887	c/ tryag	More info

FIG. 12-64

		6565 7250		
SfiI	1	4956	ggccnnnn/nggcc	More info
Sfr274I	1	5025	c/ tcgag	More info
SfuI	3	1603 1988 2423	tt/cgaa	More info
SmaI	1	4036	ccc/ggg	More info
SnaBI	1	666	tac/gta	More info
SpeI	1	326	a/ ctagt	More info
SphI	4	2930 4355 4750 4823	gcatg/c	More info
SseBI	3	3446 3546 5002	agg/cct	More info
SspBI	2	270 3471	t/ gtaca	More info
SspI	6	179 226 3571 4164 7128 7681	aat/att	More info
SstI	1	894	gagct/c	More info
StuI	3	3446 3546 5002	agg/cct	More info
StyI	13	686 1950 2226 3109 3324 3424 3547 3600 4077 4456 4574 4910 5003	c/ cwwgg	More info
Tth111I	1	3674	gacn/nngtc	More info
Van91I	6	1445 1482 1775 1796 2644 4587	ccannnn/ntgg	More info
VneI	2	5745 6991	g/ tgcac	More info
VspI	4	334 5202 5261 6496	at/ taat	More info
XbaI	1	3811	t/ ctaga	More info
XcmI	2	1948 2897	ccannnnn/nnnntgg	More info

FIG. 12-65

XhoI	1	5025	c/ tcgag	More info
XhoII	12	932 2400 2634 3409 3992 4030 6072 6083 6169 6181 6949 6966	r/ gatcy	More info
XmaI	1	4034	c/ ccggg	More info
XmaIII	1	925	c/ ggccg	More info
XmnI	5	1107 2481 3506 3906 6923	gaann/nnttc	More info
Zsp2I	5	3703 3850 4357 4752 4825	atgca/t	More info

The following endonucleases were selected but don't cut this sequence:
AccI, AccIII, AfeI, AflII, Aor51HI, AscI, BbeI, BfrI, BsaBI, Bse8I, BseAI, BsePI, Bsh1365I, BsiMI, BsiWI, Bsp13I, Bsp68I, BspEI, BspTI, BsrBRI, BssHII, Bst1107I, Bst98I, BstEII, BstPI, Cfr42I, CpoI, CspI, Eco47III, Eco91I, Eco065I, EheI, FseI, HpaI, KasI, Kpn2I, KspI, MamI, MluI, MroI, MspCI, NarI, NheI, NruI, PacI, Pfl23II, PmeI, PpuMI, PshAI, Psp5II, PspEI, PspLI, PstNHI, RsrII, SacII, SalI, SbfI, Sfr303I, SgfI, SgrAI, SmiI, SplI, SrfI, Sse8387I, SstII, SunI, SwaI, Vha464I

FIG. 12-66

FIG. 13A
FIG. 13B
FIG. 13C
FIG. 13D
FIG. 13E

FIG. 13

cccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggg
ggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagct
gatctaatcaatattggccattagccatattattcattggttatatagcataaatcaatattggctattggccattgcatacgttgtatcca
tatcataatatgtacatttatattggctcatgtccaacattaccgccatgttgacattgattattgactagttattaatagtaatcaattacg
gggtcattagttcatagcccatatatggagttccgcgttacataacttacggtaaatggcccgcctggcgaccgcccagcgaccc
ccgcccgttgacgtcaatagtgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacg
gtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtccgccccctattgacgtcaatgacggtaaatggcccgcct
agcattatgcccagtacatgaccttacgggagtttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcg
gttttggcagtacaccaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtt
tgttttggcaccaaaatcaacgggactttccaaaatgtcgtaataaccccgccccgttgacgcaaatgggcggtaggcgtgtacg
gtgggaggtctatataagcagagctcgtttagtgaaccgtcagaattcaagcttgcggccgcagatctatcgatctgcaggatatc
(EcoRV)
acc

FIG. 13A

ATGCACAGTATGATCAGCTCAGTGGATGTGAAGTCAGAAGTTCCTGTGGCCTGGAGCCCATCTCACCTTTA
GACCTAAGGACAGACCTCAGGATGATGATGCCCGTGGTGGACCCTGTTGTCCGTGAGAAGCAATTGCAGCAG
GAATTACTTCTTATCCAGCAGCAGCAACAAATCCAGAAGCAGCTTCTGATAGCAGAGTTTCAGAAACAGCAT
GAGAACTTGACACGGCAGCACCAGGCTCAGCTTCAGGAGCATATCAAGGAACTTCTAGCCATAAAACAGCAA
CAAGAACTCCTAGAAAAGGAGCAGAAACTGGAGCAGCAGAGGCAAGAACAGGAAGTAGAGAGGCATCGCAGA
GAACAGCAGCTTCCTCCTCTCAGAGGCAAAGATAGAGGACGAGAAAGGGCAGTGGCAAGTACAGAAGTAAAG
CAGAAGCTTCAAGAGTTCCTACTGAGTAAATCAGCAACGAAAGACACTCCAACTAATGGAAAAAATCATTCC
GTGAGCCGCCATCCCAAGCTCTGGTACACGGCTGCCCACCACACATCATTGGATCAAAGCTCTCCACCCCTT
AGTGGAACATCTCCATCCTACAAGTACACATTACCAGGAGCACAAGATGCAAAGGATGATTTCCCCCTTCGA
AAAACTGCCTCTGAGCCCAACTTGAAGGTGCGGTCCAGGTTAAAACAGAAAGTGGCAGAGAGGAGAAGCAGC
CCCTTACTCAGGCGGAAGGATGGAAATGTTGTCACTTCATTCAAGAAGCGAATGTTTGAGGTGACAGAATCC
TCAGTCAGTAGCAGTTCTCCAGGCTCTGGTCCCAGTTCACCAAACAATGGGCCAACTGGAAGTGTTACTGAA
AATGAGACTTCGGTTTTGCCCCCTACCCCTCATGCCGAGCAAATGGTTTCACAGCAACGCATTCTAATTCAT
GAAGATTCCATGAACCTGCTAAGTCTTTATACCTCTCCTTCTTTGCCCAACATTACCTTGGGGCTTCCCGCA
GTGCCATCCCAGCTCAATGCTTCGAATTCACTCAAAGAAAAGCAGAAGTGTGAGACGCAGACGCTTAGGCAA
GGTGTTCCTCTGCCTGGGCAGTATGGAGGCAGCATCCCGGCATCTTCCAGCCACCCTCATGTTACTTTAGAG
GGAAAGCCACCCAACAGCAGCCACCAGGCTCTCCTGCAGCATTTATTATTGAAAGAACAAATGCGACAGCAA
AAGCTTCTTGTAGCTGGTGGAGTTCCCTTACATCCTCAGTCTCCCTTGGCAACAAAAGAGAGAATTTCACCT
GGCATTAGAGGTACCCACAAATTGCCCCGTCACAGACCCCTGAACCGAACCCAGTCTGCACCTTTGCCTCAG
AGCACGTTGGCTCAGCTGGTCATTCAACAGCAACACCAGCAATTCTTGGAGAAGCAGAAGCAATACCAGCAG
CAGATCCACATGAACAAACTGCTTTCGAAATCTATTGAACAACTGAAGCAACCAGGCAGTCACCTTGAGGAA
GCAGAGGAAGAGCTTCAGGGGGACCAGGCGATGCAGGAAGACAGAGCGCCCTCTAGTGGCAACAGCACTAGG
AGCGACAGCAGTGCTTGTGTGGATGACACACTGGGACAAGTTGGGGCTGTGAAGGTCAAGGAGGAACCAGTG
GACAGTGATGAAGATGCTCAGATCCAGGAAATGGAATCTGGGGAGCAGGCTGCTTTTATGCAACAGCCTTTC

FIG. 13B

CTGGAACCCACGCACACACGTGCGCTCTCTGTGCGCCAAGCTCCGCTGGCTGCGGTTGGCATGGATGGATTA
GAGAAACACCGTCTCGTCTCCAGGACTCACTCTTCCCCTGCTGCCTCTGTTTTACCTCACCCAGCAATGGAC
CGCCCCCTCCAGCCTGGCTCTGCAACTGGAATTGCCTATGACCCCTTGATGCTGAAACACCAGTGCGTTTGT
GGCAATTCCACCACCCACCCTGAGCATGCTGGACGAATACAGAGTATCTGGTCACGACTGCAAGAAACTGGG
CTGCTAAATAAATGTGAGCGAATTCAAGGTCGAAAAGCCAGCCTGGAGGAAATACAGCTTGTTCATTCTGAA
CATCACTCACTGTTGTATGGCACCAACCCCCTGGACGGACAGAAGCTGGACCCCAGGATACTCCTAGGTGAT
GACTCTCAAAAGTTTTTTTCCTCATTACCTTGTGGTGGACTTGGGGTGGACAGTGACACCATTTGGAATGAG
CTACACTCGTCCGGTGCTGCACGCATGGCTGTTGGCTGTGTCATCGAGCTGGCTTCCAAAGTGGCCTCAGGA
GAGCTGAAGAATGGGTTTGCTGTTGTGAGGCCCCCTGGCCATCACGCTGAAGAATCCACAGCCATGGGGTTC
TGCTTTTTTAATTCAGTTGCAATTACCGCCAAATACTTGAGAGACCAACTAAATATAAGCAAGATATTGATT
GTAGATCTGGATGTTCACCATGGAAACGGTACCCAGCAGGCCTTTTATGCTGACCCCAGCATCCTGTACATT
TCACTCCATCGCTATGATGAAGGGAACTTTTTCCCTGGCAGTGGAGCCCCAAATGAGGTTCGGTTTATTTCT
TTAGAGCCCCACTTTTATTTGTATCTTTCAGGTAATTGCATTGCA

FIG. 13C

(BamHI)ggatccggtaccagattacaaggacgacgatgacaagtagatcccgggtggcatccctgtgacccctccccagtg
cctctcctggccttggaagttgccactccagtgcccaccagccttgtcctaataaaattaagttgcatcattttgtctgactaggtgtc
ctctataatattatggggtggaggggggtggtatggagcaaggggcccaagttgggaagacaacctgtagggcctgcggggtc
tattcgggaaccaagctggagtgcagtggcacaatcttggctcactgcaatctccgcctcctgggttcaagcgattctcctgcctc
agcctcccgagttgttgggattccaggcatgcatgaccaggctcagctaatttttgtttttttggtagagacggggtttcaccatattg
gccaggctggtctccaactcctaatctcaggtgatctacccaccttggcctcccaaattgctgggattacaggcgtgaaccactgc
tcccttccctgtccttctgattttaaaataactataccagcaggaggacgtccagacacagcataggctacctgccatggcccaac
cggtgggacatttgagttgcttgcttggcactgtcctctcatgcgttgggtccactcagtagatgcctgttgaattgggtacgcggc
cagcttctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctca
attagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagca
accatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaattttt
tttatttatgcagaggccgaggccgcctcggcctctgagctattccagaagtagtgaggaggcttttttggaggcctaggcttttgc
aaaaagctcctcgaggaactgaaaaaccagaaagttaattccctatagtgagtcgtattaaattcgtaatcatggtcatagctgtttc
ctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagt
gagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggcc
aacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcg
gcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagca
aaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatca
caaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcg
tgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcac
gctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgc
cttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagc
agagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatct
gcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttt

FIG. 13D

ttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtg
gaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagtttta
aatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttc
gttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgata
ccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcct
gcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgt
tgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttac
atgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcact
catggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcatt
ctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaa
gtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgt
gcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagg
gaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcg
gatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgcgccctgt
agcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctccttt
cgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgc
tttacggcacctcgaccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgcccttt
gacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataa
gggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttac
aattt

FIG. 13E

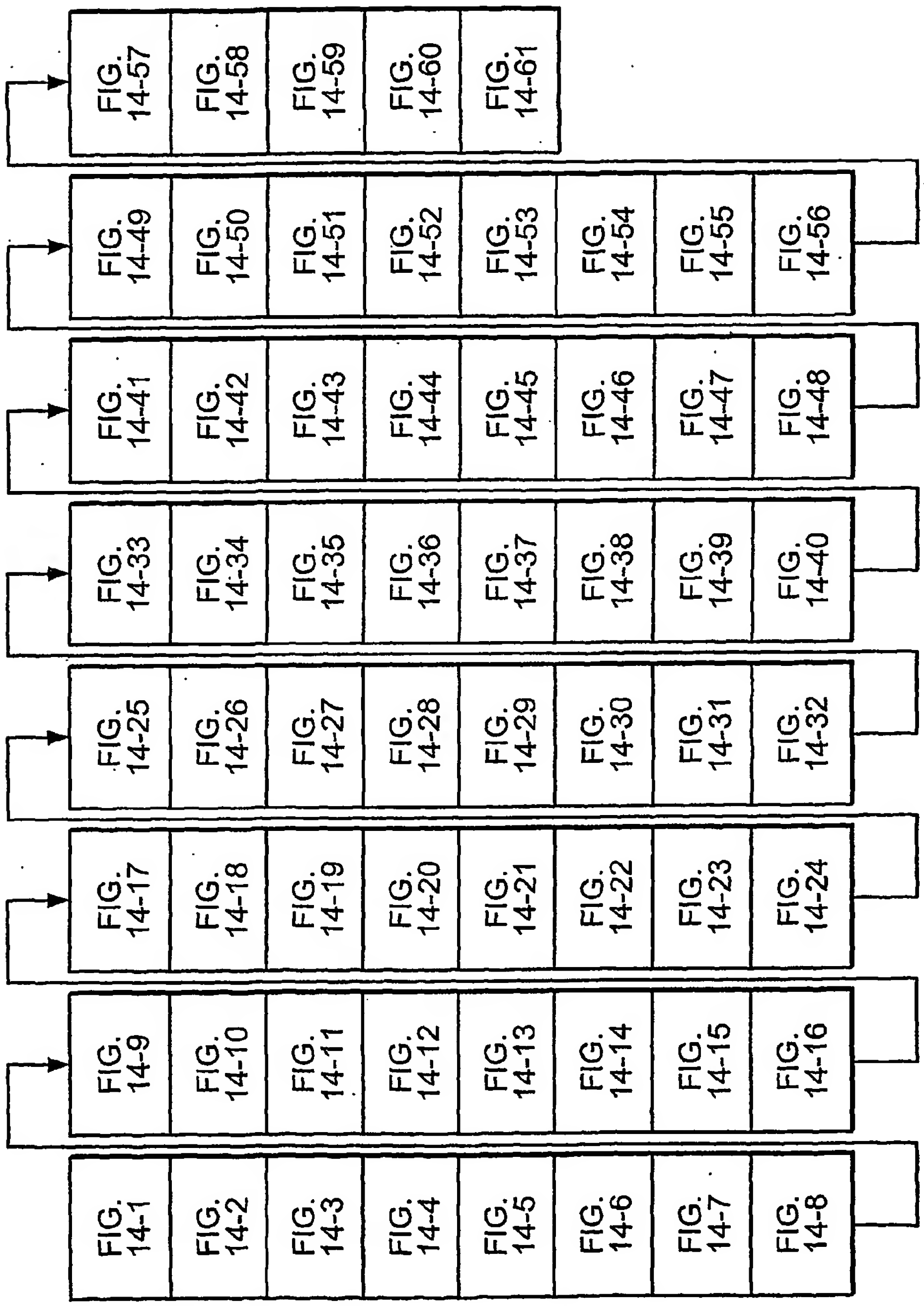


FIG. 14-1
FIG. 14-2
FIG. 14-3
FIG. 14-4
FIG. 14-5
FIG. 14-6
FIG. 14-7
FIG. 14-8
FIG. 14-9
FIG. 14-10
FIG. 14-11
FIG. 14-12
FIG. 14-13
FIG. 14-14
FIG. 14-15
FIG. 14-16
FIG. 14-17
FIG. 14-18
FIG. 14-19
FIG. 14-20
FIG. 14-21
FIG. 14-22
FIG. 14-23
FIG. 14-24
FIG. 14-25
FIG. 14-26
FIG. 14-27
FIG. 14-28
FIG. 14-29
FIG. 14-30
FIG. 14-31
FIG. 14-32
FIG. 14-33
FIG. 14-34
FIG. 14-35
FIG. 14-36
FIG. 14-37
FIG. 14-38
FIG. 14-39
FIG. 14-40
FIG. 14-41
FIG. 14-42
FIG. 14-43
FIG. 14-44
FIG. 14-45
FIG. 14-46
FIG. 14-47
FIG. 14-48
FIG. 14-49
FIG. 14-50
FIG. 14-51
FIG. 14-52
FIG. 14-53
FIG. 14-54
FIG. 14-55
FIG. 14-56
FIG. 14-57
FIG. 14-58
FIG. 14-59
FIG. 14-60
FIG. 14-61

FIG. 14

pFLAG-CMV-5b-HDAC9a

7303 base pairs

Graphic map | Table by enzyme name

```
                                        BstMCI
                   AviII                    PvuI BsiEI      EarI        MspA1I
             BglI  FspI                     BsaOI           Eam1104I    PvuII
cccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctgg
base pairs
gggtaagcggtaagtccgacgcgttgacaacccttcccgctagccacgcccggagaagcgataatgcggtcgacc
1 to 75
                   Acc16I                   BspCI           Ksp632I     NspBII
                                            Bsh1285I
                                            Ple19I
```

FIG. 14-1

```
cgaaaggggatgtgctgcaaggcgattaagttgggtaacgcccagggttttcccagtcacgacgttgtaaaacg
base pairs
gctttcccctacacgacgttccgctaattcaacccattgcgggtcccaaaagggtcagtgctgcaacattttgc
76 to 150

                                  MscI
                                CfrI
  EaeI                        SspI MluNI
acggccagtgccaagctgatctaatcaatattggccattagccatattattcattggttatatagcataaatcaa
base pairs
tgccggtcacggttcgactagattagttataaccggtaatcggtataataagtaaccaatatatcgtatttagtt
151 to 225
  CfrI                          EaeI
                                  BalI
```

FIG. 14-2

```
         MscI
         MluNI                          SspBI
 SspI   EaeI     BsrDI                  Bsp1407I
tattggctattggccattgcatacgttgtatccatatcataatatgtacatttatattggctcatgtccaacatt
base pairs
ataaccgataaccggtaacgtatgcaacataggtatagtattatacatgtaaatataaccgagtacaggttgtaa
226 to 300
        CfrI                            BsrGI
          BalI


                             VspI
        HincII        SpeI   PshBI
accgccatgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccata
base pairs
tggcggtacaactgtaactaataactgatcaataattatcattagttaatgccccagtaatcaagtatcgggtat
301 to 375
        HindII        AclNI  AsnI
                             AseI
```

FIG. 14-3

```
                                                          Hin1I
                                    BstMCI                 AcyI
                               BglI BsaOI                HincII
tatggagttccgcgttacataacttacggtaaatggcccgcctggcgaccgcccagcgacccccgcccgttgacg
base pairs
atacctcaaggcgcaatgtattgaatgccatttaccgggcggaccgctggcgggtcgctgggggcgggcaactgc
376 to 450
                                    Bsh1285I               HindII
                                    BsiEI                  Hsp92I
                                                           Msp17I

  BsaHI                                BbiII
  AatII                                Hin1I
  BbiII                                AcyI AatII
tcaatagtgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacgg
base pairs
agttatcactgcatacaagggtatcattgcggttatccctgaaaggtaactgcagttacccacctcataaatgcc
451 to 525
                                       Msp17I
                                       BsaHI
                                       Hsp92I
```

FIG. 14-4

```
                                                    BbiII
                                                    Hin1I
        BglI              NdeI                      AcyI AatII
taaactgcccacttggcagtacatcaagtgtatcatatgccaagtccgcccctattgacgtcaatgacggtaaa
base pairs
atttgacgggtgaaccgtcatgtagttcacatagtatacggttcaggcgggataactgcagttactgccattt
526 to 600
                          FauNDI                    Msp17I
                                                    BsaHI
                                                    Hsp92I

                                                          BstSNI
                                                          SnaBI
tggcccgcctagcattatgcccagtacatgaccttacgggagtttcctacttggcagtacatctacgtattagtc
base pairs
accgggcggatcgtaatacgggtcatgtactggaatgccctcaaaggatgaaccgtcatgtagatgcataatcag
601 to 675
                                                          BsaAI
                                                          Eco105I
```

FIG. 14-5

```
          NcoI Bsp19I
          StyI BstDSI
          EcoT14I
atcgctattaccatggtgatgcggttttggcagtacaccaatgggcgtggatagcggtttgactcacggggattt
base pairs
tagcgataatggtaccactacgccaaaaccgtcatgtggttacccgcacctatcgccaaactgagtgcccctaaa
676 to 750
          BssT1I
          ErhI Eco130I
          DsaI MslI

                   BbiII
                   Hin1I                AccB1I
                   AcyI AatII           BshNI
ccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgt
base pairs
ggttcagaggtggggtaactgcagttaccctcaaacaaaaccgtggttttagttgccctgaaaggttttacagca
751 to 825
                   Msp17I               BanI
                   BsaHI                Eco64I
                   Hsp92I
```

FIG. 14-6

```
                     Eco24I
  HincII                                                     EcoICRI
aataaccccgccccgttgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagca gagctcgttta
base pairs
ttattggggcggggcaactgcgtttacccgccatccgcacatgccaccctccagatatattcgt ctcgagcaaat
826 to 900
  HindII                                                     Ecl136II
                                                             Bbv12I
                                                             AspHI
                                                             Psp124BI

SacI
FriOI
SstI
BanII
BsiHKAI                   EagI XmaIII BstYI BspDI BcgI Eco32I
        AcsI              CciNI Bsh1285I BstX2I BanIII PstI          BclI
        ApoI   HindIII BstZI BstMCI MflI Bsa29I SfcI                 Ksp22I
gtgaaccgtcagaattcaagcttgcggccgcagatctatcgat ctgcaggatatcaccatgcacagtatgatcag
base pairs
cacttggcagtcttaagttcgaacgccggcgtctagatagcta gacgtcctatagtggtacgtgtcatactagtc
901 to 975
        EcoRI             EaeI Eco52I BglII BscI BspXI BstSFI        FbaI
                          CfrI EclXI BsiEI BseCI Bsu15I EcoRV
Alw21I                    NotI BsaOI XhoII ClaI Bsp106I
```

FIG. 14-7

```
                                      FriOI            CvnI        CvnI
                                     Eco24I            AocI        AocI
                                    BpmI               Bsu36I      Bsu36I
ctcagtggatgtgaagtcagaagttcctgtgggcctggagcccatctcacctttagacctaaggacagacctcag
base pairs
gagtcacctacacttcagtcttcaaggacacccggacctcgggtagagtggaaatctggattcctgtctggagtc
976 to 1050
                                    GsuI               Eco81I      Eco81I
                                      BanII            Bse21I      Bse21I

            DsaI          DrdI           MfeI             Asp700I
gatgatgatgcccgtggtggaccctgttgtccgtgagaagcaattgcagcaggaattacttcttatccagcagca
base pairs
ctactactacgggcaccacctgggacaacaggcactcttcgttaacgtcgtccttaatgaagaataggtcgtcgt
1051 to 1125
            BstDSI                       MunI             XmnI
```

FIG. 14-8

```
                          AlwNI
gcaacaaatccagaagcagcttctgatagcagagtttcagaaacagcatgagaacttgacacggcagcaccaggc
base pairs
cgttgtttaggtcttcgtcgaagactatcgtctcaaagtctttgtcgtactcttgaactgtgccgtcgtggtccg
1126 to 1200
```

```
BlpI
CelII     Eco57I                                                  EcoNI     AlwNI
tcagcttcaggagcatatcaaggaacttctagccataaaacagcaacaagaactcctagaaaggagcagaaact
base pairs
agtcgaagtcctcgtatagttccttgaagatcggtattttgtcgttgttcttgaggatctttcctcgtctttga
1201 to 1275
Bsp1720I
Bpu1102I
```

FIG. 14-9

```
    BpmI                                                          BseRI
ggagcagcagaggcaagaacaggaagtagagaggcatcgcagagaacagcagcttcctcctctcagaggcaaaga
base pairs
cctcgtcgtctccgttcttgtccttcatctctccgtagcgtctcttgtcgtcgaaggaggagagtctccgtttct
1276 to 1350
    GsuI                                                           EcoNI

                                          HindIII
tagaggacgagaaagggcagtggcaagtacagaagtaaagcag aagcttcaagagttcctactgagtaaatcagc
base pairs
atctcctgctctttcccgtcaccgttcatgtcttcatttcgtc ttcgaagttctcaaggatgactcatttagtcg
1351 to 1425
```

FIG. 14-10

```
           Van91I                                      Van91I
           AccB7I                                      AccB7I
aacgaaagacactccaactaatggaaaaatcattccgtgagccgccatcccaagctctggtacacggctgccca
base pairs
ttgctttctgtgaggttgattacctttttagtaaggcactcggcggtagggttcgagaccatgtgccgacgggt
1426 to 1500
           Esp1396I                                    Esp1396I
           PflMI                                       PflMI


ccacacatcattggatcaaagctctccacccctagtggaacatctccatcctacaagtacacattaccaggagc
base pairs
ggtgtgtagtaacctagtttcgagaggtggggaatcaccttgtagaggtaggatgttcatgtgtaatggtcctcg
1501 to 1575
```

FIG. 14-11

```
                         BstBI
Alw21I                   Bpu14I             FriOI
AspHI                    Csp45I             Eco24I
acaagatgcaaaggatgatttccccttcgaaaactgcctctgagcccaacttgaaggtgcggtccaggttaaa
base pairs
tgttctacgtttcctactaaaggggaagcttttgacggagactcgggttgaacttccacgccaggtccaattt
1576 to 1650
BsiHKAI                  SfuI Bsp119I       BanII
Bbv12I                   NspV
                         LspI
```

```
                     BseRI        EcoNI
acagaaagtggcagagaggagaagcagccccttactcaggcggaaggatggaaatgttgtcacttcattcaagaa
base pairs
tgtctttcaccgtctctcctcttcgtcggggaatgagtccgccttcctacctttacaacagtgaagtaagttctt
1651 to 1725
```

FIG. 14-12

Van91I AccB7I BpmI PflMI Van91I AccB7I

gcgaatgtttgaggtgacagaatcctcagtcagtagcagttctccaggctctggtcccagttcaccaaacaatgg
base pairs
cgcttacaaactccactgtcttaggagtcagtcatcgtcaagaggtccgagaccagggtcaagtggtttgttacc
1726 to 1800

GsuI

Esp1396I PflMI AlwNI Esp1396I

gccaactggaagtgttactgaaaatgagacttcggttttgcccctacccctcatgccgagcaaatggtttcaca
base pairs
cggttgaccttcacaatgacttttactctgaagccaaaacggggatggggagtacggctcgtttaccaaagtgt
1801 to 1875

FIG. 14-13

```
          BsaMI
          Mva1269I                BspMI                          XcmI
gcaacgcattctaattcatgaagattccatgaacctgctaagtctttatacctctccttctttgcccaacattac
base pairs
cgttgcgtaagattaagtacttctaaggtacttggacgattcagaaatatggagaggaagaaacgggttgtaatg
1876 to 1950
          BsmI RcaI
               BspHI


ErhI                                      BstBI AcsI
BssT1I                                    Bpu14I
                                          Csp45I
                                                                      Esp3I
cttggggcttcccgcagtgccatcccagctcaatgcttc gaattcactcaaagaaaagcagaagtgtgagacgca
base pairs
gaaccccgaagggcgtcacggtagggtcgagttacgaag cttaagtgagtttcttttcgtcttcacactctgcgt
1951 to 2025
EcoT14I                                   SfuI Bsp119I
                                                                      BsmBI
StyI                                      NspV ApoI
Eco130I                                   LspI EcoRI
```

FIG. 14-14

MslI
gacgcttaggcaaggtgttcctctgcctgggcagtatggaggcagcatcccggcatcttccagccaccctcatgt
base pairs
ctgcgaatccgttccacaaggagacggacccgtcatacctccgtcgtagggccgtagaaggtcggtgggagtaca
2026 to 2100

PstI
SfcI
tactttagagggaaagccacccaacagcagccaccaggctctc ctgcagcatttattattgaaagaacaaatgcg
base pairs
atgaaatctccctttcggtgggttgtcgtcggtggtccgagag gacgtcgtaaataataactttcttgtttacgc
2101 to 2175
BstSFI

FIG. 14-15

```
                                                                   Eco130I
                                                                   StyI
        HindIII                                                    EcoT14I             ApoI
acagcaaaagcttcttgtagctggtggagttcccttacatcctcagtctcccttggcaacaaaagagagaatttc
base pairs
tgtcgttttcgaagaacatcgaccacctcaagggaatgtaggagtcagagggaaccgttgttttctctcttaaag
2176 to 2250
                                                                   BssT1I              AcsI
                                                                   ErhI

              Asp718I
              Acc65I
              BshNI                                                                BsgI
acctggcattagaggtacccacaaattgccccgtcacagacccctgaaccgaacccagtctgcacctttgcctca
base pairs
tggaccgtaatctccatgggtgtttaacggggcagtgtctggggacttggcttgggtcagacgtggaaacggagt
2251 to 2325
              BanI KpnI
              AccB1I
              Eco64I
```

FIG. 14-16

```
              Bpu1102I
     Alw21I Bsp1720I
     AspHI  CelII
gagcacgttggctcagctggtcattcaacagcaacaccagcaattcttggagaagcagaagcaataccagcagca
base pairs
ctcgtgcaaccgagtcgaccagtaagttgtcgttgtggtcgttaagaacctcttcgtcttcgttatggtcgtcgt
2326 to 2400
     BsiHKAI    PvuII
     Bbv12I BlpI MspA1I
                NspBII

                       BstBI
MflI                   Bpu14I
XhoII                  Csp45I                 Eco57I
gatccacatgaacaaactgctttcgaaatctattgaacaactgaagcaaccaggcagtcaccttgaggaagcaga
base pairs
ctaggtgtacttgtttgacgaaagctttagataacttgttgacttcgttggtccgtcagtggaactccttcgtct
2401 to 2475
BstYI                  SfuI Bsp119I
BstX2I                 NspV
                       LspI
```

FIG. 14-17

```
      EarI
      Eam1104I                                  Bbv16II
     Asp700I                                    BbsI   Bsp143II
ggaagagcttcaggggaccaggcgatgcaggaagacagagcgccctctagtggcaacagcactaggagcgacag
base pairs
ccttctcgaagtcccctggtccgctacgtccttctgtctcgcgggagatcaccgttgtcgtgatcctcgctgtc
2476 to 2550
      XmnI   Eco57I                             BpiI   HaeII
      Ksp632I                                   BpuAI  BstH2I
       SapI

      BcgI
cagtgcttgtgtggatgacacactgggacaagttggggctgtgaaggtcaaggaggaaccagtggacagtgatga
base pairs
gtcacgaacacacctactgtgtgaccctgttcaaccccgacacttccagttcctccttggtcacctgtcactact
2551 to 2625
```

FIG. 14-18

```
        MflI       Van91I
        XhoII      AccB7I
agatgctcagatccaggaaatggaatctggggagcaggctgcttttatgcaacagcctttcctggaacccacgca
base pairs
tctacgagtctaggtcctttaccttagacccctcgtccgacgaaaatacgttgtcggaaaggaccttgggtgcgt
2626 to 2700
        BstYI      Esp1396I
        BstX2I     PflMI


    PmaCI
    PmlI
  AflIII                        NspBII
                                                                           Esp3I
cacacgtgcgctctctgtgcgccaagctccgctggctgcggttggcatggatggattagagaaacaccgtctcgt
base pairs
gtgtgcacgcgagagacacgcggttcgaggcgaccgacgccaaccgtacctacctaatctctttgtggcagagca
2701 to 2775
   MslI Eco72I                  MspA1I
                                                                           BsmBI
    BsaAI
    BbrPI
```

FIG. 14-19

```
              EarI
     BpmI     Eam1104I                                   BsrDI     BpmI
ctccaggactcactcttcccctgctgcctctgttttacctcacccagcaatggaccgcccctccagcctggctc
base pairs
gaggtcctgagtgagaaggggacgacggagacaaaatggagtgggtcgttacctggcgggggaggtcggaccgag
2776 to 2850
     GsuI     Ksp632I                                              GsuI

                                              XcmI
tgcaactggaattgcctatgaccccttgatgctgaaacaccagtgcgtttgtggcaattccaccacccaccctga
base pairs
acgttgaccttaacggatactggggaactacgactttgtggtcacgcaaacaccgttaaggtggtgggtgggact
2851 to 2925
```

FIG. 14-20

```
     SphI                                                              AcsI
     BbuI                                                              ApoI
gcatgctggacgaatacagagtatctggtcacgactgcaagaaactgggctgctaaataaatgtgagc gaattca
base pairs
cgtacgacctgcttatgtctcatagaccagtgctgacgttctttgacccgacgatttatttacactcg cttaagt
2926 to 3000
     PaeI                                                              EcoRI
     NspI
```

```
                      BpmI                                        AccB1I
                                                                   BshNI
aggtcgaaaagccagcctggaggaaatacagcttgttcattctgaacatcactcactgttgtatggcaccaaccc
base pairs
tccagcttttcggtcggacctcctttatgtcgaacaagtaagacttgtagtgagtgacaacataccgtggttggg
3001 to 3075
                      GsuI                                         BanI
                                                                  Eco64I
```

FIG. 14-21

```
                               ErhI
                               StyI Eco130I
BstXI           AlwNI          EcoT14I
cctggacggacagaagctggaccccaggatactcctaggtgatgactctcaaagtttttcctcattaccttg
base pairs
ggacctgcctgtcttcgacctggggtcctatgaggatccactactgagagtttcaaaaaggagtaatggaac
3076 to 3150
                               BssT1I
                               AvrII
                               BlnI

                                                   BsaWI        BsgI
tggtggacttggggtggacagtgacaccatttggaatgagctacactcgtccggtgctgcacgcatggctgttgg
base pairs
accacctgaaccccacctgtcactgtggtaaaccttactcgatgtgagcaggccacgacgtgcgtaccgacaacc
3151 to 3225
```

FIG. 14-22

```
                                    CvnI
                                    AocI                                        CfrI
                                    Bsu36I         Eco57I               DraII   EaeI
ctgtgtcatcgagctggcttccaaagtggcctcaggagagctgaagaatgggtttgctgttgtgaggccccctgg
base pairs
gacacagtagctcgaccgaaggtttcaccggagtcctctcgacttcttacccaaacgacaacactccggggacc
3226 to 3300
                                    Eco81I                             EcoO109I
                                    Bse21I


MscI                          ErhI Eco130I
                              BssT1I BstXI
                 Eco57I     MslI DsaI
ccatcacgctgaagaatccacagccatggggttctgctttttaattcagttgcaattaccgccaaatacttgag
base pairs
ggtagtgcgacttcttaggtgtcggtaccccaagacgaaaaattaagtcaacgttaatggcggtttatgaactc
3301 to 3375
MluNI                         EcoT14I
BalI                          StyI BstDSI
                              NcoI Bsp19I
```

FIG. 14-23

```
                                BstX2I           NcoI Bsp19I Asp718I    SseBI
                                 BstYI           StyI BstDSI AccB1I
Eco147I
    BsaI                         XhoII           EcoT14I   BshNI        StuI
agaccaactaaatataagcaagatattgattgtagatctggatgttcaccatggaaacggtacccagcaggcctt
base pairs
tctggttgatttatattcgttctataactaacatctagacctacaagtggtacctttgccatgggtcgtccggaa
3376 to 3450
    Eco31I                       BglII           BssT1I    BanI KpnI    AatI
                                 MflI            ErhI Eco130I Eco64I    Pme55I
                                                 DsaI      Acc65I


                 SspBI
                 Bsp1407I             MslI           Asp700I
ttatgctgaccccagcatcctgtacatttcactccatcgctatgatgaagggaacttttccctggcagtggagc
base pairs
aatacgactggggtcgtaggacatgtaaagtgaggtagcgatactacttcccttgaaaagggaccgtcacctcg
3451 to 3525
                 BsrGI                               XmnI
```

FIG. 14-24

```
 FriOI                          FriOI                                 BstYI
 Eco24I                         Eco24I                                XhoII
                                                                    BsrDI
cccaaatgaggttcggtttatttctttagagccccactttatttgtatctttcaggtaattgcattgca ggatc
base pairs
gggtttactccaagccaaataaagaaatctcggggtgaaaataaacatagaaagtccattaacgtaacgt cctag
3526 to 3600
 BanII                          BanII                                 BamHI
                                                                      BstI
                                                                      MflI

Acc65I                             AvaI BcoI
  BanI Eco64I                     MflI Eco88I PspALI
BstX2I Asp718I                    XhoII Cfr9I SmaI MslI
cggtaccagattacaaggacgacgatgacaagtagat cccgggtggcatccctgtgacccctccccagtgcctct
base pairs
gccatggtctaatgttcctgctgctactgttcatcta gggcccaccgtagggacactggggaggggtcacggaga
3601 to 3675
  BshNI                           BstYI Ama87I
 BsaWI KpnI                       BstX2I BsoBI
  AccB1I                              XmaI PspAI
```

FIG. 14-25

```
    Eco130I
    StyI            GsuI
    EcoT14I       MslI
cctggccttggaagttgccactccagtgcccaccagccttgtcctaataaaattaagttgcatcattttgtctga
base pairs
ggaccggaaccttcaacggtgaggtcacgggtggtcggaacaggattattttaattcaacgtagtaaaacagact
3676 to 3750
    BssT1I          BpmI
    ErhI

                                                         Eco24I
                                                                      SfcI
   Eam1105I                                           DraII BanII     Bbv16II
   AspEI       SspI                                   PspOMI FriOI    BbsI
ctaggtgtcctctataatattatggggtggagggggtggtatggagcaaggggcccaagttggaagacaacct
base pairs
gatccacaggagatattataataccccacctcccccaccatacctcgttccccgggttcaacccttctgttgga
3751 to 3825
   EclHKI                                             Bsp120I         BpiI
   AhdI                                               EcoO109I        BpuAI
                                                          ApaI        BstSFI
```

FIG. 14-26

```
    DraII                                        BpmI BsgI
gtagggcctgcggggtctattcgggaaccaagctggagtgcagtggcacaatcttggctcactgcaatctccgcc
base pairs
catcccggacgccccagataagcccttggttcgacctcacgtcaccgtgttagaaccgagtgacgttagaggcgg
3826 to 3900
    EcoO109I                                     GsuI

                                            BcoI                           NspI       BlpI
                                           Ama87I                          PaeI Mph1103I
                                  BcgI      AvaI                         Ppu10I EcoT22I
tcctgggttcaagcgattctcctgcctcagcctcccgagttgttgggattccaggcatgcatgaccaggctcagc
base pairs
aggacccaagttcgctaagaggacggagtcggagggctcaacaaccctaaggtccgtacgtactggtccgagtcg
3901 to 3975
                                           Eco88I                          BbuI Zsp2I CelII
                                           BsoBI                           SphI       Bsp172
                                                                             NsiI     Bpu11
```

FIG. 14-27

```
                         MscI
                         MluNI
          Esp3I         EaeI         BsaI
taattttgtttttttggtagagacggggtttcaccatattggccaggctggtctccaactcctaatctcaggtg
base pairs
attaaaacaaaaaaccatctctgccccaaagtggtataaccggtccgaccagaggttgaggattagagtccac
3976 to 4050
          BsmBI         CfrI         Eco31I
                         BalI
 0I
02I

     Eco130I
     StyI
     EcoT14I     BstXI
atctacccaccttggcctcccaaattgctgggattacaggcgtgaaccactgctcccttccctgtccttctgatt
base pairs
tagatgggtggaaccggagggtttaacgaccctaatgtccgcacttggtgacgagggaagggacaggaagactaa
4051 to 4125
     BssT1I
     ErhI
```

FIG. 14-28

```
                     BbiII                       NcoI Eco130I BsrFI PflMI
                     Hin1I                       StyI DsaI AgeI Bse118I
 DraI                AcyI AatII                  EcoT14I   BsaWI AccB7I
ttaaaataactataccagcaggaggacgtccagacacagcataggctacctgccatggcccaaccggtgggacat
base pairs
aatttattgatatggtcgtcctcctgcaggtctgtgtcgtatccgatggacggtaccgggttggccaccctgta
4126 to 4200
                     Msp17I                      BssT1I    BssAI Esp1396I
                     BsaHI                       ErhI BstDSI PinAI Van91I
                     Hsp92I                      BspMI Bsp19I Cfr10I

                                                                      EaeI
ttgagttgcttgcttggcactgtcctctcatgcgttgggtccactcagtagatgcctgttgaattgggtacgcgg
base pairs
aactcaacgaacgaaccgtgacaggagagtacgcaacccaggtgagtcatctacggacaacttaacccatgcgcc
4201 to 4275
                                                                      CfrI
```

FIG. 14-29

```
      AlwNI
ccagcttctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaag
base pairs
ggtcgaagacaccttacacacagtcaatcccacacctttcaggggtccgaggggtcgtccgtcttcatacgtttc
4276 to 4350



    NspI
    PaeI Mph1103I
  Ppu10I EcoT22I        SexAI
catgcatctcaattagtcagcaaccaggtgtggaaaagtccccaggctccccagcaggcagaagtatgcaaagca
base pairs
gtacgtagagttaatcagtcgttggtccacaccttttcaggggtccgaggggtcgtccgtcttcatacgtttcgt
4351 to 4425
    BbuI Zsp2I
    SphI
      NsiI
```

FIG. 14-30

```
  NspI
  PaeI Mph1103I
Ppu10I EcoT22I
tgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccg
base pairs
acgtagagttaatcagtcgttggtatcagggcggggattgaggcgggtagggcggggattgaggcgggtcaaggc
4426 to 4500
  BbuI Zsp2I
  SphI
    NsiI

              NcoI Bsp19I
              StyI BstDSI
              EcoT14I                               BglI
cccattctccgccccatggctgactaattttttttatttatgcagaggccgaggccgcctcggcctctgagctat
base pairs
gggtaagaggcggggtaccgactgattaaaaaaaataaatacgtctccggctccggcggagccggagactcgata
4501 to 4575
              BssT1I                                SfiI
              ErhI Eco130I
              DsaI
```

FIG. 14-31

```
                        SseBI AvrII            Ama87I
                        Eco147I BlnI           Eco88I BseRI
              BseRI     StuI BssT1I            AvaI BsoBI
tccagaagtagtgaggaggctttttggaggcctaggcttttgcaaaagctc ctcgaggaactgaaaaaccaga
base pairs
aggtcttcatcactcctccgaaaaacctccggatccgaaaacgttttcgag gagctccttgacttttggtct
4576 to 4650
                        AatI StyI              XhoI BcoI
                        Pme55I ErhI            Sfr274I
                         EcoT14I Eco130I       PaeR7I


          SfcI               ApoI
aagttaattccctatagtgagtcgtattaaattcgtaatcatggtcatagctgtttcctgtgtgaaattgttatc
base pairs
ttcaattaagggatatcactcagcataatttaagcattagtaccagtatcgacaaaggacacactttaacaatag
4651 to 4725
          BstSFI             AcsI
```

FIG. 14-32

```
     AccBSI                                                   AccB1I
     BsrBI                                                    BshNI
cgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaac
base pairs
gcgagtgttaaggtgtgttgtatgctcggccttcgtatttcacatttcggaccccacggattactcactcgattg
4726 to 4800
     BstD102I                                                 BanI
                                                              Eco64I

                                                                     VspI
     VspI                                                      MspA1I
     PshBI                                                      PvuII PshBI
                                                                           EaeI
tcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcg
base pairs
agtgtaattaacgcaacgcgagtgacgggcgaaaggtcagccctttggacagcacggtcgacgtaattacttagc
4801 to 4875
     AsnI                                                       NspBII
                                                                           CfrI
     AseI                                                             AsnI
                                                                      AseI
```

FIG. 14-33

```
                                             Eam1104I
                                        BstH2I
                                        Bsp143II
gccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcgg
base pairs
cggttgcgcgcccctctccgccaaacgcataacccgcgagaaggcgaaggagcgagtgactgagcgacgcgagcc
4876 to 4950
                                        HaeII EarI
                                             SapI
                                             Ksp632I

 BstMCI               AccBSI
 BsaOI                BsrBI
tcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacg
base pairs
agcaagccgacgccgctcgccatagtcgagtgagtttccgccattatgccaataggtgtcttagtcccctattgc
4951 to 5025
 Bsh1285I             BstD102I
 BsiEI
```

FIG. 14-34

```
      NspI
  BspLU11I
caggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttcc
base pairs
gtcctttcttgtacactcgttttccggtcgttttccggtccttggcatttttccggcgcaacgaccgcaaaaagg
5026 to 5100
     AflIII
```

```
                      DrdI
ataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactat
base pairs
tatccgaggcggggggactgctcgtagtgtttttagctgcgagttcagtctccaccgctttgggctgtcctgata
5101 to 5175
```

FIG. 14-35

```
                                BsiI                                  BsaWI
aaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacc
base pairs
tttctatggtccgcaaagggggaccttcgagggagcacgcgagaggacaaggctgggacggcgaatggcctatgg
5176 to 5250
                                BssSI

                             BstH2I
                             Bsp143II        SfcI
tgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtagg
base pairs
acaggcggaaagagggaagcccttcgcaccgcgaaagagttacgagtgcgacatccatagagtcaagccacatcc
5251 to 5325
                             HaeII           BstSFI
```

FIG. 14-36

```
                BsiHKAI                   NspBII
            Alw44I                       BstMCI
            VneI Bbv12I                  BsaOI        BsaWI
tcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatc
base pairs
agcaagcgaggttcgacccgacacacgtgcttggggggcaagtcgggctggcgacgcggaataggccattgatag
5326 to 5400
            ApaLI                        Bsh1285I
                AspHI                    BsiEI
                Alw21I                    MspA1I



                                          AlwNI
gtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcga
base pairs
cagaactcaggttgggccattctgtgctgaatagcggtgaccgtcgtcggtgaccattgtcctaatcgtctcgct
5401 to 5475
```

FIG. 14-37

```
               SfcI
ggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggta
base pairs
ccatacatccgccacgatgtctcaagaacttcaccaccggattgatgccgatgtgatcttcttgtcataaaccat
5476 to 5550
               BstSFI


                Eco57I
                                                                  NspBII
tctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctg
base pairs
agacgcgagacgacttcggtcaatggaagcctttttctcaaccatcgagaactaggccgtttgtttggtggcgac
5551 to 5625
                                                                  MspA1I
```

FIG. 14-38

```
                                                                  MflI        MflI
                                                                  XhoII       XhoII
gtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatct
base pairs
catcgccaccaaaaaaacaaacgttcgtcgtctaatgcgcgtcttttttttcctagagttcttctaggaaactaga
5626 to 5700
                                                                  BstYI       BstYI
                                                                  BstX2I      BstX2I


                                                                      RcaI            MflI
                                                                                     XhoII
tttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaagga
base pairs
aaagatgccccagactgcgagtcaccttgcttttgagtgcaattccctaaaaccagtactctaatagtttttcct
5701 to 5775
                                                                      BspHI          BstYI
                                                                                     BstX2I
```

FIG. 14-39

```
          MflI
          XhoII    DraI                 DraI
tcttcacctagatcctttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctg
base pairs
agaagtggatctaggaaatttaatttttacttcaaaatttagttagatttcatatatactcatttgaaccagac
5776 to 5850
          BstYI
          BstX2I

                          AccB1I
                          BshNI
acagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgac
base pairs
tgtcaatggttacgaattagtcactccgtggatagagtcgctagacagataaagcaagtaggtatcaacggactg
5851 to 5925
                          BanI
                          Eco64I
```

FIG. 14-40

```
   Eam1105I
   AspEI                                                        BsrDI
tccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacc
base pairs
aggggcagcacatctattgatgctatgccctcccgaatggtagaccggggtcacgacgttactatggcgctctgg
5926 to 6000
   EclHKI
   AhdI


        Cfr10I
BsaI    BssAI     BpmI                           BglI
cacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaa
base pairs
gtgcgagtggccgaggtctaaatagtcgttatttggtcggtcggccttcccggctcgcgtcttcaccaggacgtt
6001 to 6075
Eco31I  BsrFI     GsuI
        Bse118I
```

FIG. 14-41

```
                         VspI
                         PshBI
ctttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgc
base pairs
gaaataggcggaggtaggtcagataattaacaacggcccttcgatctcattcatcaagcggtcaattatcaaacg
6076 to 6150
                         AsnI
                         AseI
```

```
AviII              BstSFI
FspI               SfcI        MslI                                   BsaWI
gcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggtt
base pairs
cgttgcaacaacggtaacgatgtccgtagcaccacagtgcgagcagcaaaccataccgaagtaagtcgaggccaa
6151 to 6225
Acc16I             BsrDI
    Psp1406I
```

FIG. 14-42

BsiEI
PvuI
BstMCI
BsaOI
cccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcg
base pairs
gggttgctagttccgctcaatgtactagggggtacaacacgttttttcgccaatcgaggaagccaggaggctagc
6226 to 6300
BspCI
Bsh1285I
Ple19I

EaeI MslI
ttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgc
base pairs
aacagtcttcattcaaccggcgtcacaatagtgagtaccaataccgtcgtgacgtattaagagaatgacagtacg
6301 to 6375
CfrI

FIG. 14-43

```
                   Acc113I
                                                   BstMCI
                   Eco255I                          BsaOI
catccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccga
base pairs
gtaggcattctacgaaaagacactgaccactcatgagttggttcagtaagactcttatcacatacgccgctggct
6376 to 6450
                   ScaI
                                                  Bsh1285I
                                                   BsiEI

       BbiII
       Hin1I                                    Alw21I
  BcgI AcyI                               DraI  AspHI
gttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaa
base pairs
caacgagaacgggccgcagttatgccctattatggcgcggtgtatcgtcttgaaattttcacgagtagtaacctt
6451 to 6525
       Msp17I                                    BsiHKAI
       BsaHI                                     Bbv12I
       Hsp92I
```

FIG. 14-44

```
 XmnI                        MflI             MflI                   BssSI
 Psp1406I                    XhoII   NspBII   XhoII                Alw44I
                                                                   VneI
aacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcac
base pairs
ttgcaagaagccccgcttttgagagttcctagaatggcgacaactctaggtcaagctacattgggtgagcacgtg
6526 to 6600
 Asp700I                     BstYI   MspA1I   BstYI                ApaLI
                             BstX2I           BstX2I                BsiI
                                                                    AspHI
```

```
 Bbv12I
BsiHKAI        Eco57I
ccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaa
base pairs
ggttgactagaagtcgtagaaaatgaaagtggtcgcaaagacccactcgtttttgtccttccgttttacggcgtt
6601 to 6675
 Alw21I
```

FIG. 14-45

```
                    MslI                          EarI
                                                  Eam1104I  SspI
aaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatc
base pairs
ttttcccttattcccgctgtgcctttacaacttatgagtatgagaaggaaaaagttataataacttcgtaaatag
6676 to 6750
                                                  Ksp632I
```

```
                    AccBSI
              RcaI  BsrBI
agggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacat
base pairs
tcccaataacagagtactcgcctatgtataaacttacataaatctttttatttgtttatccccaaggcgcgtgta
6751 to 6825
              BspHI BstD102I
```

FIG. 14-46

```
                             SfcI
ttccccgaaaagtgccacctgacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcg
base pairs
aaggggcttttcacggtggactgcgcgggacatcgccgcgtaattcgcgccgcccacaccaccaatgcgcgtcgc
6826 to 6900
                             BstSFI


                                          AccBSI                            BsrFI
                      BstH2I    HaeII  BstD102I                             BssAI
                      Bsp143II          BsrBI                               MroNI
tgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccg
base pairs
actggcgatgtgaacggtcgcgggatcgcgggcgaggaaagcgaaagaagggaaggaaagagcggtgcaagcggc
6901 to 6975                                                               NgoAIV
                      HaeII     Bsp143II                                    NgoMI
                                                                           Bse118I
                                BstH2I
```

FIG. 14-47

NaeI AccB1I BshNI

gctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtgctttacggcacctcgacccca
base pairs
cgaaaggggcagttcgagatttagccccgtagggaaatcccaaggctaaatcacgaaatgccgtggagctggggt
6976 to 7050

Cfr10I BanI Eco64I

BsaAI DrdI

aaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttgg
base pairs
tttttgaactaatcccactaccaagtgcatcacccggtagcgggactatctgccaaaaagcgggaaactgcaacc
7051 to 7125

DraIII

FIG. 14-48

```
agtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcggtctattctttg
base pairs
tcaggtgcaagaaattatcacctgagaacaaggtttgaccttgttgtgagttgggatagagccagataagaaaac
7126 to 7200

                                                                      ApoI       ApoI
atttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaatt
base pairs
taaatattccctaaaacggctaaagccggataaccaattttttactcgactaaattgtttttaaattgcgcttaa
7201 to 7275
                                                                      AcsI       AcsI

          SspI  Psp1406I
ttaacaaaatattaaacgtttacaattt  base pairs
aattgttttataatttgcaaatgttaaa  7276 to 7303
```

FIG. 14-49

Table by Enzyme Name

Enzyme name	No. cuts	Positions of sites	Recognition sequence	
AatI	2	3446 4606	agg/cct	More info
AatII	5	451 504 587 773 4154	gacgt/c	More info
Acc113I	1	6408	agt/act	More info
Acc16I	2	21 6150	tgc/gca	More info
Acc65I	3	2264 3434 3602	g/ gtacc	More info
AccB1I	8	791 2264 3065 3434 3602 4779 5876 7036	g/ gyrcc	More info
AccB7I	6	1445 1482 1775 1796 2644 4191	ccannnn/ntgg	More info
AccBSI	4	4730 4971 6772 6936	gagcgg	More info
AclNI	1	326	a/ ctagt	More info
AcsI	7	912 1990 2244 2994 4679 7260 7271	r/ aatty	More info
AcyI	6	448 501 584 770 4151 6465	gr/cgyc	More info
AflIII	2	2702 5035	a/ crygt	More info
AgeI	1	4188	a/ ccggt	More info
AhdI	2	3754 5928	gacnnn/nngtc	More info
Alw21I	6	894 1576 2330 5353 6514 6599	gwgcw/c	More info
Alw44I	2	5349 6595	g/ tgcac	More info

FIG. 14-50

AlwNI	6	1147 1273 1775 3091 4282 5451	cagnnn/ctg	More info
Ama87I	3	3638 3934 4629	c/ ycgrg	More info
AocI	3	1034 1046 3256	cc/ tnagg	More info
ApaI	1	3806	gggcc/c	More info
ApaLI	2	5349 6595	g/ tgcac	More info
ApoI	7	912 1990 2244 2994 4679 7260 7271	r/ aatty	More info
AseI	4	334 4806 4865 6100	at/ taat	More info
AsnI	4	334 4806 4865 6100	at/ taat	More info
Asp700I	4	1107 2481 3506 6527	gaann/nnttc	More info
Asp718I	3	2264 3434 3602	g/ gtacc	More info
AspEI	2	3754 5928	gacnnn/nngtc	More info
AspHI	6	894 1576 2330 5353 6514 6599	gwgcw/c	More info
AvaI	3	3638 3934 4629	c/ ycgrg	More info
AviII	2	21 6150	tgc/gca	More info
AvrII	2	3109 4607	c/ ctagg	More info
BalI	4	184 238 3300 4018	tgg/cca	More info
BamHI	1	3596	g/ gatcc	More info
BanI	8	791 2264 3065 3434 3602 4779 5876 7036	g/ gyrcc	More info
BanII	6	894 1017 1623 3526 3558 3806	grgcy/c	More info
BanIII	1	939	at/ cgat	More info
BbiII	6	448 501 584 770 4151 6465	gr/cgyc	More info

FIG. 14-51

BbrPI	1	2705	cac/gtg	More info
BbsI	2	2512 3820	gaagac	More info
BbuI	4	2930 3959 4354 4427	gcatg/c	More info
Bbv12I	6	894 1576 2330 5353 6514 6599	gwgcw/c	More info
Bbv16II	2	2512 3820	gaagac	More info
BcgI	4	941 2556 3925 6455	cgannnnnntgc	More info
BclI	1	969	t/ gatca	More info
BcoI	3	3638 3934 4629	c/ ycgrg	More info
BglI	5	14 417 538 4560 6048	gccnnnn/nggc	More info
BglII	2	932 3409	a/ gatct	More info
BlnI	2	3109 4607	c/ ctagg	More info
BlpI	3	1200 2337 3970	gc/tnagc	More info
BpiI	2	2512 3820	gaagac	More info
BpmI	9	1015 1279 1772 2781 2842 3022 3701 3863 6018	ctggag	More info
Bpu1102I	3	1200 2337 3970	gc/tnagc	More info
Bpu14I	3	1603 1988 2423	tt/cgaa	More info
BpuAI	2	2512 3820	gaagac	More info
Bsa29I	1	939	at/ cgat	More info
BsaAI	3	666 2705 7077	yac/gtr	More info
BsaHI	6	448 501 584 770 4151 6465	gr/cgyc	More info

FIG. 14-52

BsaI	3	3380 4031 6000	ggtctc	More info
BsaMI	1	1886	gaatgc	More info
BsaOI	7	42 424 928 4951 5375 6298 6447	cgry/cg	More info
BsaWI	6	3200 3599 4188 5241 5388 6219	w/ ccggw	More info
BscI	1	939	at/ cgat	More info
Bse118I	3	4188 6008 6972	r/ ccggy	More info
Bse21I	3	1034 1046 3256	cc/ tnagg	More info
BseCI	1	939	at/ cgat	More info
BseRI	4	1337 1671 4593 4631	gaggag	More info
BsgI	3	2315 3212 3868	gtgcag	More info
Bsh1285I	7	42 424 928 4951 5375 6298 6447	cgry/cg	More info
BshNI	8	791 2264 3065 3434 3602 4779 5876 7036	g/ gyrcc	More info
BsiEI	7	42 424 928 4951 5375 6298 6447	cgry/cg	More info
BsiHKAI	6	894 1576 2330 5353 6514 6599	gwgcw/c	More info
BsiI	2	5213 6597	ctcgtg	More info
BsmBI	3	2023 2773 4001	cgtctc	More info
BsmI	1	1886	gaatgc	More info
BsoBI	3	3638 3934 4629	c/ ycgrg	More info
Bsp106I	1	939	at/ cgat	More info
Bsp119I	3	1603 1988 2423	tt/cgaa	More info
Bsp120I	1	3802	g/ ggccc	More info
Bsp1407I	2	270 3471	t/ gtaca	More info

FIG. 14-53

Bsp143II	5	2519 4913 5283 6922 6930	rgcgc/y	More info
Bsp1720I	3	1200 2337 3970	gc/tnagc	More info
Bsp19I	5	686 3324 3424 4178 4514	c/ catgg	More info
BspCI	2	42 6298	cgat/cg	More info
BspDI	1	939	at/ cgat	More info
BspHI	3	1891 5755 6763	t/ catga	More info
BspLU11I	1	5035	a/ catgt	More info
BspMI	2	1913 4178	acctgc	More info
BspXI	1	939	at/ cgat	More info
BsrBI	4	4730 4971 6772 6936	gagcgg	More info
BsrDI	5	245 2827 3594 5987 6169	gcaatg	More info
BsrFI	3	4188 6008 6972	r/ ccggy	More info
BsrGI	2	270 3471	t/ gtaca	More info
BssAI	3	4188 6008 6972	r/ ccggy	More info
BssSI	2	5213 6597	ctcgtg	More info
BssT1I	11	686 1950 2226 3109 3324 3424 3681 4060 4178 4514 4607	c/ cwwgg	More info
BstBI	3	1603 1988 2423	tt/cgaa	More info
BstD102I	4	4730 4971 6772 6936	gagcgg	More info
BstDSI	6	686 1062 3324 3424 4178 4514	c/ crygg	More info
BstH2I	5	2519 4913 5283 6922 6930	rgcgc/y	More info
BstI	1	3596	g/ gatcc	More info

FIG. 14-54

BstMCI	7	42 424 928 4951 5375 6298 6447	cgry/cg	More info
BstSFI	8	944 2144 3824 4662 5300 5491 6169 6854	c/ tryag	More info
BstSNI	1	666	tac/gta	More info
BstX2I	12	932 2400 2634 3409 3596 3634 5676 5687 5773 5785 6553 6570	r/ gatcy	More info
BstXI	3	3076 3325 4077	ccannnnn/ntgg	More info
BstYI	12	932 2400 2634 3409 3596 3634 5676 5687 5773 5785 6553 6570	r/ gatcy	More info
BstZI	1	925	c/ ggccg	More info
Bsu15I	1	939	at/ cgat	More info
Bsu36I	3	1034 1046 3256	cc/ tnagg	More info
CciNI	1	925	gc/ggccgc	More info
CelII	3	1200 2337 3970	gc/tnagc	More info
Cfr10I	3	4188 6008 6972	r/ ccggy	More info
Cfr9I	1	3638	c/ ccggg	More info
CfrI	9	152 182 236 925 3298 4016 4273 4874 6316	y/ ggccr	More info
ClaI	1	939	at/ cgat	More info
Csp45I	3	1603 1988 2423	tt/cgaa	More info
CvnI	3	1034 1046 3256	cc/ tnagg	More info

FIG. 14-55

DraI	4	4127 5794 5813 6505	ttt/aaa	More info
DraII	3	3291 3802 3829	rg/gnccy	More info
DraIII	1	7080	cacnnn/gtg	More info
DrdI	3	1076 5143 7124	gacnnnn/nngtc	More info
DsaI	6	686 1062 3324 3424 4178 4514	c/ crygg	More info
EaeI	9	152 182 236 925 3298 4016 4273 4874 6316	y/ ggccr	More info
EagI	1	925	c/ ggccg	More info
Eam1104I	5	58 2482 2793 4918 6722	ctcttc	More info
Eam1105I	2	3754 5928	gacnnn/nngtc	More info
EarI	5	58 2482 2793 4918 6722	ctcttc	More info
Ecl136II	1	892	gag/ ctc	More info
EclHKI	2	3754 5928	gacnnn/nngtc	More info
EclXI	1	925	c/ ggccg	More info
Eco105I	1	666	tac/gta	More info
Eco130I	11	686 1950 2226 3109 3324 3424 3681 4060 4178 4514 4607	c/ cwwgg	More info
Eco147I	2	3446 4606	agg/cct	More info
Eco24I	6	894 1017 1623 3526 3558 3806	grgcy/c	More info
Eco255I	1	6408	agt/act	More info
Eco31I	3	3380 4031 6000	ggtctc	More info
Eco32I	1	952	gat/ atc	More info
Eco52I	1	925	c/ ggccg	More info

FIG. 14-56

Eco57I	7	1210 2446 2488 3271 3314 5567 6615	ctgaag	More info
Eco64I	8	791 2264 3065 3434 3602 4779 5876 7036	g/ gyrcc	More info
Eco72I	1	2705	cac/gtg	More info
Eco81I	3	1034 1046 3256	cc/ tnagg	More info
Eco88I	3	3638 3934 4629	c/ ycgrg	More info
EcoICRI	1	892	gag/ ctc	More info
EcoNI	3	1259 1338 1684	cctnn/nnnagg	More info
EcoO109I	3	3291 3802 3829	rg/gnccy	More info
EcoRI	3	912 1990 2994	g/ aattc	More info
EcoRV	1	952	gat/ atc	More info
EcoT14I	11	686 1950 2226 3109 3324 3424 3681 4060 4178 4514 4607	c/ cwwgg	More info
EcoT22I	3	3961 4356 4429	atgca/t	More info
ErhI	11	686 1950 2226 3109 3324 3424 3681 4060 4178 4514 4607	c/ cwwgg	More info
Esp1396I	6	1445 1482 1775 1796 2644 4191	ccannnn/ntgg	More info
Esp3I	3	2023 2773 4001	cgtctc	More info
FauNDI	1	560	ca/ tatg	More info
FbaI	1	969	t/ gatca	More info
FriOI	6	894 1017 1623 3526 3558 3806	grgcy/c	More info
FspI	2	21 6150	tgc/gca	More info

FIG. 14-57

GsuI	9	1015 1279 1772 2781 2842 3022 3701 3863 6018	ctggag	More info
HaeII	5	2519 4913 5283 6922 6930	rgcgc/y	More info
Hin1I	6	448 501 584 770 4151 6465	gr/cgyc	More info
HincII	3	311 446 842	gty/rac	More info
HindII	3	311 446 842	gty/rac	More info
HindIII	3	918 1394 2183	a/ agctt	More info
Hsp92I	6	448 501 584 770 4151 6465	gr/cgyc	More info
KpnI	3	2268 3438 3606	ggtac/c	More info
Ksp22I	1	969	t/ gatca	More info
Ksp632I	5	58 2482 2793 4918 6722	ctcttc	More info
LspI	3	1603 1988 2423	tt/cgaa	More info
MfeI	1	1091	c/ aattg	More info
MflI	12	932 2400 2634 3409 3596 3634 5676 5687 5773 5785 6553 6570	r/ gatcy	More info
MluNI	4	184 238 3300 4018	tgg/cca	More info
Mph1103I	3	3961 4356 4429	atgca/t	More info
MroNI	1	6972	g/ ccggc	More info
MscI	4	184 238 3300 4018	tgg/cca	More info
MslI	10	691 2094 2703 3323 3489 3651 3698 6180 6339 6698	caynn/nnrtg	More info
Msp17I	6	448 501 584 770 4151 6465	gr/cgyc	More info
MspA1I	7	71 2341 2731 4859 5377 5622 6563	cmg/ckg	More info

FIG. 14-58

MunI	1	1091	c/ aattg	More info
Mva1269I	1	1886	gaatgc	More info
NaeI	1	6974	gcc/ggc	More info
NcoI	5	686 3324 3424 4178 4514	c/ catgg	More info
NdeI	1	560	ca/ tatg	More info
NgoAIV	1	6972	g/ ccggc	More info
NgoMI	1	6972	g/ ccggc	More info
NotI	1	925	gc/ggccgc	More info
NsiI	3	3961 4356 4429	atgca/t	More info
NspBII	7	71 2341 2731 4859 5377 5622 6563	cmg/ckg	More info
NspI	5	2930 3959 4354 4427 5039	rcatg/y	More info
NspV	3	1603 1988 2423	tt/cgaa	More info
PaeI	4	2930 3959 4354 4427	gcatg/c	More info
PaeR7I	1	4629	c/ tcgag	More info
PflMI	6	1445 1482 1775 1796 2644 4191	ccannnn/ntgg	More info
PinAI	1	4188	a/ ccggt	More info
Ple19I	2	42 6298	cgat/cg	More info
PmaCI	1	2705	cac/gtg	More info
Pme55I	2	3446 4606	agg/cct	More info
PmlI	1	2705	cac/gtg	More info
Ppu10I	3	3957 4352 4425	a/ tgcat	More info
PshBI	4	334 4806 4865 6100	at/ taat	More info
Psp124BI	1	894	gagct/c	More info
Psp1406I	3	6154 6527 7291	aa/cgtt	More info

FIG. 14-59

PspAI	1	3638	c/ ccggg	More info
PspALI	1	3640	ccc/ggg	More info
PspOMI	1	3802	g/ ggccc	More info
PstI	2	948 2148	ctgca/g	More info
PvuI	2	42 6298	cgat/cg	More info
PvuII	3	71 2341 4859	cag/ctg	More info
RcaI	3	1891 5755 6763	t/ catga	More info
SacI	1	894	gagct/c	More info
SapI	2	2483 4918	gctcttc	More info
ScaI	1	6408	agt/act	More info
SexAI	1	4373	a/ ccwggt	More info
SfcI	8	944 2144 3824 4662 5300 5491 6169 6854	c/ tryag	More info
SfiI	1	4560	ggccnnnn/nggcc	More info
Sfr274I	1	4629	c/ tcgag	More info
SfuI	3	1603 1988 2423	tt/cgaa	More info
SmaI	1	3640	ccc/ggg	More info
SnaBI	1	666	tac/gta	More info
SpeI	1	326	a/ ctagt	More info
SphI	4	2930 3959 4354 4427	gcatg/c	More info
SseBI	2	3446 4606	agg/cct	More info
SspBI	2	270 3471	t/ gtaca	More info
SspI	5	179 226 3768 6732 7285	aat/att	More info
SstI	1	894	gagct/c	More info

FIG. 14-60

StuI	2	3446 4606	agg/cct	More info
StyI	11	686 1950 2226 3109 3324 3424 3681 4060 4178 4514 4607	c/ cwwgg	More info
Van91I	6	1445 1482 1775 1796 2644 4191	ccannnn/ntgg	More info
VneI	2	5349 6595	g/ tgcac	More info
VspI	4	334 4806 4865 6100	at/ taat	More info
XcmI	2	1948 2897	ccannnnn/nnnntgg	More info
XhoI	1	4629	c/ tcgag	More info
XhoII	12	932 2400 2634 3409 3596 3634 5676 5687 5773 5785 6553 6570	r/gatcy	More info
XmaI	1	3638	c/ ccggg	More info
XmaIII	1	925	c/ ggccg	More info
XmnI	4	1107 2481 3506 6527	gaann/nnttc	More info
Zsp2I	3	3961 4356 4429	atgca/t	More info

The following endonucleases were selected but don't cut this sequence:
AccI, AccIII, AfeI, AflII, Aor51HI, AscI, AspI, AtsI, BbeI, BfrI, BsaBI, Bse8I, BseAI, BsePI, Bsh1365I, BsiMI, BsiWI, Bsp13I, Bsp68I, BspEI, BspTI, BsrBRI, BssHII, Bst1107I, Bst98I, BstEII, BstPI, Cfr42I, CpoI, CspI, Eco47III, Eco91I, EcoO65I, EheI, FseI, HpaI, KasI, Kpn2I, KspI, MamI, MluI, MroI, MspCI, NarI, NheI, NruI, PacI, Pfl23II, PmeI, PpuMI, PshAI, Psp5II, PspEI, PspLI, PstNHI, RsrII, SacII, SalI, SbfI, Sfr303I, SgfI, SgrAI, SmiI, SplI, SrfI, Sse8387I, SstII, SunI, SwaI, Tth111I, Vha464I, XbaI

FIG. 14-61

SEQUENCE LISTING

<110> Sloan-Kettering Institute for Cancer Research
Richon, Victoria
Zhou, Xianbo
Rifkind, Richard A.
Marks, Paul A.

<120> HDAC9 Polypeptides and Polynucleotides
and Uses Thereof

<130> 3254.1000005

<150> 60/298,173
<151> 2001-06-14

<150> 60/311,686
<151> 2001-08-10

<150> 60/316,995
<151> 2001-09-04

<160> 22

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 3186
<212> DNA
<213> Homo sapiens

<400> 1
ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60
tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tggggtggct 120
ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtggatgtg 180
aagtcagaag ttcctgtggg cctggagccc atctcacctt tagacctaag gacagacctc 240
aggatgatga tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300
cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360
cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420
ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480
caagaacagg aagtagagag gcatcgcaga gaacagcagc ttcctcctct cagaggcaaa 540
gatagaggac gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600
ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660
cgccatccca agctctggta cacggctgcc caccacacat cattggatca aagctctcca 720
ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaag 780
gatgatttcc cccttcgaaa aactgcctct gagcccaact tgaaggtgcg gtccaggtta 840
aaacagaaag tggcagagag gagaagcagc cccttactca ggcggaagga tggaaatgtt 900
gtcacttcat tcaagaagcg aatgtttgag gtgacagaat cctcagtcag tagcagttct 960
ccaggctctg gtcccagttc accaaacaat gggccaactg gaagtgttac tgaaaatgag 1020
acttcggttt tgccccctac ccctcatgcc gagcaaatgg tttcacagca acgcattcta 1080
attcatgaag attccatgaa cctgctaagt ctttatacct ctccttcttt gcccaacatt 1140
accttggggc ttcccgcagt gccatcccag ctcaatgctt cgaattcact caaagaaaag 1200
cagaagtgtg agacgcagac gcttaggcaa ggtgttcctc tgcctgggca gtatggaggc 1260
agcatcccgg catcttccag ccaccctcat gttactttag agggaaagcc acccaacagc 1320
agccaccagg ctctcctgca gcatttatta ttgaaagaac aaatgcgaca gcaaaagctt 1380
cttgtagctg gtggagttcc cttacatcct cagtctccct tggcaacaaa agagagaatt 1440
tcacctggca ttagaggtac ccacaaattg ccccgtcaca gacccctgaa ccgaacccag 1500
tctgcacctt tgcctcagag cacgttggct cagctggtca ttcaacagca acaccagcaa 1560
ttcttggaga agcagaagca ataccagcag cagatccaca tgaacaaact gctttcgaaa 1620
tctattgaac aactgaagca accaggcagt caccttgagg aagcagagga agagcttcag 1680
ggggaccagg cgatgcagga agacagagcg ccctctagtg gcaacagcac taggagcgac 1740

agcagtgctt gtgtggatga cacactggga caagttgggg ctgtgaaggt caaggaggaa 1800
ccagtggaca gtgatgaaga tgctcagatc caggaaatgg aatctgggga gcaggctgct 1860
tttatgcaac agcctttcct ggaacccacg cacacacgtg cgctctctgt gcgccaagct 1920
ccgctggctg cggttggcat ggatggatta gagaaacacc gtctcgtctc caggactcac 1980
tcttcccctg ctgcctctgt tttacctcac ccagcaatgg accgccccct ccagcctggc 2040
tctgcaactg gaattgccta tgaccccttg atgctgaaac accagtgcgt ttgtggcaat 2100
tccaccaccc accctgagca tgctggacga atacagagta tctggtcacg actgcaagaa 2160
actgggctgc taaataaatg tgagcgaatt caaggtcgaa aagccagcct ggaggaaata 2220
cagcttgttc attctgaaca tcactcactg ttgtatggca ccaacccccct ggacggacag 2280
aagctggacc ccaggatact cctaggtgat gactctcaaa agttttttc ctcattacct 2340
tgtggtggac ttggggtgga cagtgacacc atttggaatg agctacactc gtccggtgct 2400
gcacgcatgg ctgttggctg tgtcatcgag ctggcttcca aagtggcctc aggagagctg 2460
aagaatgggt ttgctgttgt gaggcccct ggccatcacg ctgaagaatc cacagccatg 2520
gggttctgct tttttaattc agttgcaatt accgccaaat acttgagaga ccaactaaat 2580
ataagcaaga tattgattgt agatctggat gttcaccatg gaaacggtac ccagcaggcc 2640
ttttatgctg accccagcat cctgtacatt tcactccatc gctatgatga agggaacttt 2700
ttccctggca gtggagcccc aaatgaggtt ggaacaggcc ttggagaagg gtacaatata 2760
aatattgcct ggacaggtgg ccttgatcct cccatgggag atgttgagta ccttgaagca 2820
ttcaggacca tcgtgaagcc tgtggccaaa gagtttgatc cagacatggt cttagtatct 2880
gctggatttg atgcattgga aggccacacc cctcctctag gagggtacaa agtgacggca 2940
aaatgttttg gtcatttgac gaagcaattg atgacattgg ctgatggacg tgtggtgttg 3000
gctctagaag gaggacatga tctcacagcc atctgtgatg catcagaagc ctgtgtaaat 3060
gcccttctag gaaatgagct ggagccactt gcagaagata ttctccacca aagcccgaat 3120
atgaatgctg ttatttcttt acagaagatc attgaaattc aaagtatgtc tttaaagttc 3180
tcttaa 3186

<210> 2
<211> 1011
<212> PRT
<213> Homo sapiens

<400> 2
Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
1 5 10 15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
20 25 30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
35 40 45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
50 55 60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65 70 75 80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
85 90 95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
100 105 110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
115 120 125
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
130 135 140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145 150 155 160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
165 170 175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
180 185 190
Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
195 200 205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Ala Ser Glu Pro Asn Leu
210 215 220
Lys Val Arg Ser Arg Leu Lys Gln Lys Val Ala Glu Arg Arg Ser Ser
225 230 235 240

Pro Leu Leu Arg Arg Lys Asp Gly Asn Val Val Thr Ser Phe Lys Lys
245 250 255
Arg Met Phe Glu Val Thr Glu Ser Ser Val Ser Ser Ser Ser Pro Gly
260 265 270
Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly Ser Val Thr Glu
275 280 285
Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala Glu Gln Met Val
290 295 300
Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met Asn Leu Leu Ser
305 310 315 320
Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu Gly Leu Pro Ala
325 330 335
Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys Glu Lys Gln Lys
340 345 350
Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu Pro Gly Gln Tyr
355 360 365
Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His Val Thr Leu Glu
370 375 380
Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu Gln His Leu Leu
385 390 395 400
Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val Ala Gly Gly Val
405 410 415
Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu Arg Ile Ser Pro
420 425 430
Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg Pro Leu Asn Arg
435 440 445
Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala Gln Leu Val Ile
450 455 460
Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys Gln Tyr Gln Gln
465 470 475 480
Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile Glu Gln Leu Lys
485 490 495
Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu Leu Gln Gly Asp
500 505 510
Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly Asn Ser Thr Arg
515 520 525
Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly Gln Val Gly Ala
530 535 540
Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu Asp Ala Gln Ile
545 550 555 560
Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met Gln Gln Pro Phe
565 570 575
Leu Glu Pro Thr His Thr Arg Ala Leu Ser Val Arg Gln Ala Pro Leu
580 585 590
Ala Ala Val Gly Met Asp Gly Leu Glu Lys His Arg Leu Val Ser Arg
595 600 605
Thr His Ser Ser Pro Ala Ala Ser Val Leu Pro His Pro Ala Met Asp
610 615 620
Arg Pro Leu Gln Pro Gly Ser Ala Thr Gly Ile Ala Tyr Asp Pro Leu
625 630 635 640
Met Leu Lys His Gln Cys Val Cys Gly Asn Ser Thr Thr His Pro Glu
645 650 655
His Ala Gly Arg Ile Gln Ser Ile Trp Ser Arg Leu Gln Glu Thr Gly
660 665 670
Leu Leu Asn Lys Cys Glu Arg Ile Gln Gly Arg Lys Ala Ser Leu Glu
675 680 685
Glu Ile Gln Leu Val His Ser Glu His His Ser Leu Leu Tyr Gly Thr
690 695 700
Asn Pro Leu Asp Gly Gln Lys Leu Asp Pro Arg Ile Leu Leu Gly Asp
705 710 715 720
Asp Ser Gln Lys Phe Phe Ser Ser Leu Pro Cys Gly Gly Leu Gly Val
725 730 735

Asp Ser Asp Thr Ile Trp Asn Glu Leu His Ser Ser Gly Ala Ala Arg
740 745 750
Met Ala Val Gly Cys Val Ile Glu Leu Ala Ser Lys Val Ala Ser Gly
755 760 765
Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro Gly His His Ala
770 775 780
Glu Glu Ser Thr Ala Met Gly Phe Cys Phe Phe Asn Ser Val Ala Ile
785 790 795 800
Thr Ala Lys Tyr Leu Arg Asp Gln Leu Asn Ile Ser Lys Ile Leu Ile
805 810 815
Val Asp Leu Asp Val His His Gly Asn Gly Thr Gln Gln Ala Phe Tyr
820 825 830
Ala Asp Pro Ser Ile Leu Tyr Ile Ser Leu His Arg Tyr Asp Glu Gly
835 840 845
Asn Phe Phe Pro Gly Ser Gly Ala Pro Asn Glu Val Gly Thr Gly Leu
850 855 860
Gly Glu Gly Tyr Asn Ile Asn Ile Ala Trp Thr Gly Gly Leu Asp Pro
865 870 875 880
Pro Met Gly Asp Val Glu Tyr Leu Glu Ala Phe Arg Thr Ile Val Lys
885 890 895
Pro Val Ala Lys Glu Phe Asp Pro Asp Met Val Leu Val Ser Ala Gly
900 905 910
Phe Asp Ala Leu Glu Gly His Thr Pro Pro Leu Gly Gly Tyr Lys Val
915 920 925
Thr Ala Lys Cys Phe Gly His Leu Thr Lys Gln Leu Met Thr Leu Ala
930 935 940
Asp Gly Arg Val Val Leu Ala Leu Glu Gly Gly His Asp Leu Thr Ala
945 950 955 960
Ile Cys Asp Ala Ser Glu Ala Cys Val Asn Ala Leu Leu Gly Asn Glu
965 970 975
Leu Glu Pro Leu Ala Glu Asp Ile Leu His Gln Ser Pro Asn Met Asn
980 985 990
Ala Val Ile Ser Leu Gln Lys Ile Ile Glu Ile Gln Ser Met Ser Leu
995 1000 1005
Lys Phe Ser
1010

<210> 3
<211> 3499
<212> DNA
<213> Homo sapiens

<400> 3
ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60
tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tggggtggct 120
ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtggatgtg 180
aagtcagaag ttcctgtggg cctggagccc atctcacctt tagacctaag gacagacctc 240
aggatgatga tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300
cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360
cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420
ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480
caagaacagg aagtagagag gcatcgcaga gaacagcagc ttcctcctct cagaggcaaa 540
gatagaggac gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600
ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660
cgccatccca agctctggta cacggctgcc caccacacat cattggatca aagctctcca 720
ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaag 780
gatgatttcc cccttcgaaa aactgcctct gagcccaact tgaaggtgcg gtccaggtta 840
aaacagaaag tggcagagag gagaagcagc cccttactca ggcggaagga tggaaatgtt 900
gtcacttcat tcaagaagcg aatgtttgag gtgacagaat cctcagtcag tagcagttct 960
ccaggctctg gtcccagttc accaaacaat gggccaactg gaagtgttac tgaaaatgag 1020
acttcggttt tgccccctac ccctcatgcc gagcaaatgg tttcacagca acgcattcta 1080

attcatgaag attccatgaa cctgctaagt ctttatacct ctccttcttt gcccaacatt 1140
accttggggc ttcccgcagt gccatcccag ctcaatgctt cgaattcact caaagaaaag 1200
cagaagtgtg agacgcagac gcttaggcaa ggtgttcctc tgcctgggca gtatggaggc 1260
agcatcccgg catcttccag ccaccctcat gttactttag agggaaagcc acccaacagc 1320
agccaccagg ctctcctgca gcatttatta ttgaaagaac aaatgcgaca gcaaaagctt 1380
cttgtagctg gtggagttcc cttacatcct cagtctccct tggcaacaaa agagagaatt 1440
tcacctggca ttagaggtac ccacaaattg ccccgtcaca gacccctgaa ccgaacccag 1500
tctgcacctt tgcctcagag cacgttggct cagctggtca ttcaacagca acaccagcaa 1560
ttcttggaga agcagaagca ataccagcag cagatccaca tgaacaaact gctttcgaaa 1620
tctattgaac aactgaagca accaggcagt caccttgagg aagcagagga agagcttcag 1680
ggggaccagg cgatgcagga agacagagcg ccctctagtg gcaacagcac taggagcgac 1740
agcagtgctt gtgtggatga cacactggga caagttgggg ctgtgaaggt caaggaggaa 1800
ccagtggaca gtgatgaaga tgctcagatc caggaaatgg aatctgggga gcaggctgct 1860
tttatgcaac agcctttcct ggaacccacg cacacacgtg cgctctctgt gcgccaagct 1920
ccgctggctg cggttggcat ggatggatta gagaaacacc gtctcgtctc caggactcac 1980
tcttcccctg ctgcctctgt tttacctcac ccagcaatgg accgccccct ccagcctggc 2040
tctgcaactg gaattgccta tgaccccttg atgctgaaac accagtgcgt ttgtggcaat 2100
tccaccaccc accctgagca tgctggacga atacagagta tctggtcacg actgcaagaa 2160
actgggctgc taaataaatg tgagcgaatt caaggtcgaa aagccagcct ggaggaaata 2220
cagcttgttc attctgaaca tcactcactg ttgtatggca ccaacccccct ggacggacag 2280
aagctggacc ccaggatact cctaggtgat gactctcaaa agtttttttc ctcattacct 2340
tgtggtggac ttggggtgga cagtgacacc atttggaatg agctacactc gtccggtgct 2400
gcacgcatgg ctgttggctg tgtcatcgag ctggcttcca aagtggcctc aggagagctg 2460
aagaatgggt ttgctgttgt gaggcccccct ggccatcacg ctgaagaatc cacagccatg 2520
ggttctgct tttttaattc agttgcaatt accgccaaat acttgagaga ccaactaaat 2580
ataagcaaga tattgattgt agatctggat gttcaccatg gaaacggtac ccagcaggcc 2640
ttttatgctg accccagcat cctgtacatt tcactccatc gctatgatga agggaacttt 2700
ttccctggca gtggagcccc aaatgaggtt cggtttattt ctttagagcc ccacttttat 2760
ttgtatcttt caggtaattg cattgcatga ttaccccctaa ttttcttgtc ctttgctggt 2820
gttttaaatt acacgagatt actgaattgt cccatgggac caagaaccag tgcagaacaa 2880
gtgcataacc cagagcactg tttgtcaggg aaggttgggc tgatttgatg tgttgtttga 2940
tgtttatttc aagagctccc atgtgcttgt tttcctctct tcttgctttc ttccatttgc 3000
tctcttctct gcccaccgtg gtgtgtcttt ctcttcccag gttggaacag gccttggaga 3060
agggtacaat ataaatattg cctggacagg tggccttgat cctcccatgg gagatgttga 3120
gtaccttgaa gcattcagga ccatcgtgaa gcctgtggcc aaagagtttg atccagacat 3180
ggtcttagta tctgctggat ttgatgcatt ggaaggccac acccctcctc taggagggta 3240
caaagtgacg gcaaaatgtt ttggtcattt gacgaagcaa ttgatgacat tggctgatgg 3300
acgtgtggtg ttggctctag aaggaggaca tgatctcaca gccatctgtg atgcatcaga 3360
agcctgtgta aatgcccttc taggaaatga gctggagcca cttgcagaag atattctcca 3420
ccaaagcccg aatatgaatg ctgttatttc tttacagaag atcattgaaa ttcaaagtat 3480
gtctttaaag ttctcttaa 3499

<210> 4
<211> 879
<212> PRT
<213> Homo sapiens

<400> 4
Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
1 5 10 15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
20 25 30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
35 40 45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
50 55 60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65 70 75 80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
85 90 95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
100 105 110

Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
115 120 125
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
130 135 140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145 150 155 160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
165 170 175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
180 185 190
Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
195 200 205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Ala Ser Glu Pro Asn Leu
210 215 220
Lys Val Arg Ser Arg Leu Lys Gln Lys Val Ala Glu Arg Arg Ser Ser
225 230 235 240
Pro Leu Leu Arg Arg Lys Asp Gly Asn Val Val Thr Ser Phe Lys Lys
245 250 255
Arg Met Phe Glu Val Thr Glu Ser Ser Val Ser Ser Ser Ser Pro Gly
260 265 270
Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly Ser Val Thr Glu
275 280 285
Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala Glu Gln Met Val
290 295 300
Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met Asn Leu Leu Ser
305 310 315 320
Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu Gly Leu Pro Ala
325 330 335
Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys Glu Lys Gln Lys
340 345 350
Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu Pro Gly Gln Tyr
355 360 365
Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His Val Thr Leu Glu
370 375 380
Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu Gln His Leu Leu
385 390 395 400
Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val Ala Gly Gly Val
405 410 415
Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu Arg Ile Ser Pro
420 425 430
Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg Pro Leu Asn Arg
435 440 445
Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala Gln Leu Val Ile
450 455 460
Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys Gln Tyr Gln Gln
465 470 475 480
Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile Glu Gln Leu Lys
485 490 495
Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu Leu Gln Gly Asp
500 505 510
Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly Asn Ser Thr Arg
515 520 525
Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly Gln Val Gly Ala
530 535 540
Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu Asp Ala Gln Ile
545 550 555 560
Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met Gln Gln Pro Phe
565 570 575
Leu Glu Pro Thr His Thr Arg Ala Leu Ser Val Arg Gln Ala Pro Leu
580 585 590
Ala Ala Val Gly Met Asp Gly Leu Glu Lys His Arg Leu Val Ser Arg
595 600 605

Thr His Ser Ser Pro Ala Ala Ser Val Leu Pro His Pro Ala Met Asp
610 615 620
Arg Pro Leu Gln Pro Gly Ser Ala Thr Gly Ile Ala Tyr Asp Pro Leu
625 630 635 640
Met Leu Lys His Gln Cys Val Cys Gly Asn Ser Thr Thr His Pro Glu
645 650 655
His Ala Gly Arg Ile Gln Ser Ile Trp Ser Arg Leu Gln Glu Thr Gly
660 665 670
Leu Leu Asn Lys Cys Glu Arg Ile Gln Gly Arg Lys Ala Ser Leu Glu
675 680 685
Glu Ile Gln Leu Val His Ser Glu His His Ser Leu Leu Tyr Gly Thr
690 695 700
Asn Pro Leu Asp Gly Gln Lys Leu Asp Pro Arg Ile Leu Leu Gly Asp
705 710 715 720
Asp Ser Gln Lys Phe Phe Ser Ser Leu Pro Cys Gly Gly Leu Gly Val
725 730 735
Asp Ser Asp Thr Ile Trp Asn Glu Leu His Ser Ser Gly Ala Ala Arg
740 745 750
Met Ala Val Gly Cys Val Ile Glu Leu Ala Ser Lys Val Ala Ser Gly
755 760 765
Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro Gly His His Ala
770 775 780
Glu Glu Ser Thr Ala Met Gly Phe Cys Phe Phe Asn Ser Val Ala Ile
785 790 795 800
Thr Ala Lys Tyr Leu Arg Asp Gln Leu Asn Ile Ser Lys Ile Leu Ile
805 810 815
Val Asp Leu Asp Val His His Gly Asn Gly Thr Gln Gln Ala Phe Tyr
820 825 830
Ala Asp Pro Ser Ile Leu Tyr Ile Ser Leu His Arg Tyr Asp Glu Gly
835 840 845
Asn Phe Phe Pro Gly Ser Gly Ala Pro Asn Glu Val Arg Phe Ile Ser
850 855 860
Leu Glu Pro His Phe Tyr Leu Tyr Leu Ser Gly Asn Cys Ile Ala
865 870 875

<210> 5
<211> 3054
<212> DNA
<213> Homo sapiens

<400> 5
ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60
tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tggggtggct 120
ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtggatgtg 180
aagtcagaag ttcctgtggg cctggagccc atctcacctt tagacctaag gacagacctc 240
aggatgatga tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300
cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360
cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420
ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480
caagaacagg aagtagagag gcatcgcaga gaacagcagc ttcctcctct cagaggcaaa 540
gatagaggac gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600
ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660
cgccatccca agctctggta cacggctgcc caccacacat cattggatca aagctctcca 720
ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaag 780
gatgatttcc cccttcgaaa aactgaatcc tcagtcagta gcagttctcc aggctctggt 840
cccagttcac caaacaatgg gccaactgga agtgttactg aaaatgagac ttcggttttg 900
ccccctaccc ctcatgccga gcaaatggtt tcacagcaac gcattctaat tcatgaagat 960
tccatgaacc tgctaagtct ttatacctct ccttctttgc ccaacattac cttggggctt 1020
cccgcagtgc catcccagct caatgcttcg aattcactca aagaaaagca gaagtgtgag 1080
acgcagacgc ttaggcaagg tgttcctctg cctgggcagt atggaggcag catcccggca 1140
tcttccagcc accctcatgt tactttagag ggaaagccac ccaacagcag ccaccaggct 1200

```
ctcctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct tgtagctggt 1260
ggagttccct tacatcctca gtctcccttg gcaacaaaag agagaatttc acctggcatt 1320
agaggtaccc acaaattgcc ccgtcacaga cccctgaacc gaacccagtc tgcacctttg 1380
cctcagagca cgttggctca gctggtcatt caacagcaac accagcaatt cttggagaag 1440
cagaagcaat accagcagca gatccacatg aacaaactgc tttcgaaatc tattgaacaa 1500
ctgaagcaac caggcagtca ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg 1560
atgcaggaag acagagcgcc ctctagtggc aacagcacta ggagcgacag cagtgcttgt 1620
gtggatgaca cactgggaca agttggggct gtgaaggtca aggaggaacc agtggacagt 1680
gatgaagatg ctcagatcca ggaaatggaa tctggggagc aggctgcttt tatgcaacag 1740
cctttcctgg aacccacgca cacacgtgcg ctctctgtgc gccaagctcc gctggctgcg 1800
gttggcatgg atggattaga gaaacaccgt ctcgtctcca ggactcactc ttcccctgct 1860
gcctctgttt tacctcaccc agcaatggac cgccccctcc agcctggctc tgcaactgga 1920
attgcctatg accccttgat gctgaaacac cagtgcgttt gtggcaattc caccacccac 1980
cctgagcatg ctggacgaat acagagtatc tggtcacgac tgcaagaaac tgggctgcta 2040
aataaatgtg agcgaattca aggtcgaaaa gccagcctgg aggaaataca gcttgttcat 2100
tctgaacatc actcactgtt gtatggcacc aacccccctgg acggacagaa gctggacccc 2160
aggatactcc taggtgatga ctctcaaaag ttttttttcct cattaccttg tggtggactt 2220
ggggtggaca gtgacaccat ttggaatgag ctacactcgt ccggtgctgc acgcatggct 2280
gttggctgtg tcatcgagct ggcttccaaa gtggcctcag gagagctgaa gaatgggttt 2340
gctgttgtga ggccccctgg ccatcacgct gaagaatcca cagccatggg gttctgcttt 2400
tttaattcag ttgcaattac cgccaaatac ttgagagacc aactaaatat aagcaagata 2460
ttgattgtag atctggatgt tcaccatgga aacggtaccc agcaggcctt ttatgctgac 2520
cccagcatcc tgtacatttc actccatcgc tatgatgaag ggaacttttt ccctggcagt 2580
ggagccccaa atgaggttgg aacaggcctt ggagaagggt acaatataaa tattgcctgg 2640
acaggtggcc ttgatcctcc catgggagat gttgagtacc ttgaagcatt caggaccatc 2700
gtgaagcctg tggccaaaga gtttgatcca gacatggtct tagtatctgc tggatttgat 2760
gcattggaag gccacacccc tcctctagga gggtacaaag tgacggcaaa atgttttggt 2820
catttgacga agcaattgat gacattggct gatggacgtg tggtgttggc tctagaagga 2880
ggacatgatc tcacagccat ctgtgatgca tcagaagcct gtgtaaatgc ccttctagga 3940
aatgagctgg agccacttgc agaagatatt ctccaccaaa gcccgaatat gaatgctgtt 3000
atttctttac agaagatcat tgaaattcaa agtatgtctt taaagttctc ttaa       3054
```

<210> 6
<211> 967
<212> PRT
<213> Homo sapiens

<400> 6
```
Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
 1               5                   10                  15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
            20                  25                  30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
        35                  40                  45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
    50                  55                  60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65                  70                  75                  80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
                85                  90                  95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
            100                 105                 110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
        115                 120                 125
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
    130                 135                 140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145                 150                 155                 160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
                165                 170                 175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
            180                 185                 190
```

```
Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
        195                 200                 205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Glu Ser Ser Val Ser Ser
    210                 215                 220
Ser Ser Pro Gly Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly
225                 230                 235                 240
Ser Val Thr Glu Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala
                245                 250                 255
Glu Gln Met Val Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met
            260                 265                 270
Asn Leu Leu Ser Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu
        275                 280                 285
Gly Leu Pro Ala Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys
    290                 295                 300
Glu Lys Gln Lys Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu
305                 310                 315                 320
Pro Gly Gln Tyr Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His
                325                 330                 335
Val Thr Leu Glu Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu
            340                 345                 350
Gln His Leu Leu Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val
        355                 360                 365
Ala Gly Gly Val Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu
    370                 375                 380
Arg Ile Ser Pro Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg
385                 390                 395                 400
Pro Leu Asn Arg Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala
                405                 410                 415
Gln Leu Val Ile Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys
            420                 425                 430
Gln Tyr Gln Gln Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile
        435                 440                 445
Glu Gln Leu Lys Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu
    450                 455                 460
Leu Gln Gly Asp Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly
465                 470                 475                 480
Asn Ser Thr Arg Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly
                485                 490                 495
Gln Val Gly Ala Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu
            500                 505                 510
Asp Ala Gln Ile Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met
        515                 520                 525
Gln Gln Pro Phe Leu Glu Pro Thr His Thr Arg Ala Leu Ser Val Arg
    530                 535                 540
Gln Ala Pro Leu Ala Ala Val Gly Met Asp Gly Leu Glu Lys His Arg
545                 550                 555                 560
Leu Val Ser Arg Thr His Ser Ser Pro Ala Ala Ser Val Leu Pro His
                565                 570                 575
Pro Ala Met Asp Arg Pro Leu Gln Pro Gly Ser Ala Thr Gly Ile Ala
            580                 585                 590
Tyr Asp Pro Leu Met Leu Lys His Gln Cys Val Cys Gly Asn Ser Thr
        595                 600                 605
Thr His Pro Glu His Ala Gly Arg Ile Gln Ser Ile Trp Ser Arg Leu
    610                 615                 620
Gln Glu Thr Gly Leu Leu Asn Lys Cys Glu Arg Ile Gln Gly Arg Lys
625                 630                 635                 640
Ala Ser Leu Glu Glu Ile Gln Leu Val His Ser Glu His His Ser Leu
                645                 650                 655
Leu Tyr Gly Thr Asn Pro Leu Asp Gly Gln Lys Leu Asp Pro Arg Ile
            660                 665                 670
Leu Leu Gly Asp Asp Ser Gln Lys Phe Phe Ser Ser Leu Pro Cys Gly
        675                 680                 685
```

Gly Leu Gly Val Asp Ser Asp Thr Ile Trp Asn Glu Leu His Ser Ser
    690   695   700
Gly Ala Ala Arg Met Ala Val Gly Cys Val Ile Glu Leu Ala Ser Lys
705   710   715   720
Val Ala Ser Gly Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro
                725   730   735
Gly His His Ala Glu Glu Ser Thr Ala Met Gly Phe Cys Phe Phe Asn
            740   745   750
Ser Val Ala Ile Thr Ala Lys Tyr Leu Arg Asp Gln Leu Asn Ile Ser
        755   760   765
Lys Ile Leu Ile Val Asp Leu Asp Val His His Gly Asn Gly Thr Gln
    770   775   780
Gln Ala Phe Tyr Ala Asp Pro Ser Ile Leu Tyr Ile Ser Leu His Arg
785   790   795   800
Tyr Asp Glu Gly Asn Phe Phe Pro Gly Ser Gly Ala Pro Asn Glu Val
                805   810   815
Gly Thr Gly Leu Gly Glu Gly Tyr Asn Ile Asn Ile Ala Trp Thr Gly
            820   825   830
Gly Leu Asp Pro Pro Met Gly Asp Val Glu Tyr Leu Glu Ala Phe Arg
        835   840   845
Thr Ile Val Lys Pro Val Ala Lys Glu Phe Asp Pro Asp Met Val Leu
    850   855   860
Val Ser Ala Gly Phe Asp Ala Leu Glu Gly His Thr Pro Pro Leu Gly
865   870   875   880
Gly Tyr Lys Val Thr Ala Lys Cys Phe Gly His Leu Thr Lys Gln Leu
                885   890   895
Met Thr Leu Ala Asp Gly Arg Val Val Leu Ala Leu Glu Gly Gly His
            900   905   910
Asp Leu Thr Ala Ile Cys Asp Ala Ser Glu Ala Cys Val Asn Ala Leu
        915   920   925
Leu Gly Asn Glu Leu Glu Pro Leu Ala Glu Asp Ile Leu His Gln Ser
    930   935   940
Pro Asn Met Asn Ala Val Ile Ser Leu Gln Lys Ile Ile Glu Ile Gln
945   950   955   960
Ser Met Ser Leu Lys Phe Ser
                965

<210> 7
<211> 3367
<212> DNA
<213> Homo sapiens

<400> 7
ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60
tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tggggtggct 120
ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtggatgtg 180
aagtcagaag ttcctgtggg cctggagccc atctcacctt tagacctaag gacagacctc 240
aggatgatga tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300
cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360
cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420
ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480
caagaacagg aagtagagag gcatcgcaga gaacagcagc ttcctcctct cagaggcaaa 540
gatagaggac gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600
ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660
cgccatccca agctctggta cacggctgcc caccacacat cattggatca aagctctcca 720
ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaag 780
gatgatttcc cccttcgaaa aactgaatcc tcagtcagta gcagttctcc aggctctggt 840
cccagttcac caaacaatgg gccaactgga agtgttactg aaaatgagac ttcggttttg 900
ccccctaccc ctcatgccga gcaaatggtt tcacagcaac gcattctaat tcatgaagat 960
tccatgaacc tgctaagtct ttatacctct ccttctttgc ccaacattac cttggggctt 1020
cccgcagtgc catcccagct caatgcttcg aattcactca aagaaaagca gaagtgtgag 1080

```
acgcagacgc ttaggcaagg tgttcctctg cctgggcagt atggaggcag catcccggca 1140
tcttccagcc accctcatgt tactttagag ggaaagccac ccaacagcag ccaccaggct 1200
ctcctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct tgtagctggt 1260
ggagttccct tacatcctca gtctcccttg gcaacaaaag agagaatttc acctggcatt 1320
agaggtaccc acaaattgcc ccgtcacaga cccctgaacc gaacccagtc tgcacctttg 1380
cctcagagca cgttggctca gctggtcatt caacagcaac accagcaatt cttggagaag 1440
cagaagcaat accagcagca gatccacatg aacaaactgc tttcgaaatc tattgaacaa 1500
ctgaagcaac caggcagtca ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg 1560
atgcaggaag acagagcgcc ctctagtggc aacagcacta ggagcgacag cagtgcttgt 1620
gtggatgaca cactgggaca agttggggct gtgaaggtca aggaggaacc agtggacagt 1680
gatgaagatg ctcagatcca ggaaatggaa tctggggagc aggctgcttt tatgcaacag 1740
cctttcctgg aacccacgca cacacgtgcg ctctctgtgc gccaagctcc gctggctgcg 1800
gttggcatgg atggattaga gaaacaccgt ctcgtctcca ggactcactc ttcccctgct 1860
gcctctgttt tacctcaccc agcaatggac cgcccccctcc agcctggctc tgcaactgga 1920
attgcctatg acccccttgat gctgaaacac cagtgcgttt gtggcaattc caccacccac 1980
cctgagcatg ctggacgaat acagagtatc tggtcacgac tgcaagaaac tgggctgcta 2040
aataaatgtg agcgaattca aggtcgaaaa gccagcctgg aggaaataca gcttgttcat 2100
tctgaacatc actcactgtt gtatggcacc aacccccctgg acggacagaa gctggacccc 2160
aggatactcc taggtgatga ctctcaaaag tttttttcct cattaccttg tggtggactt 2220
ggggtggaca gtgacaccat ttggaatgag ctacactcgt ccggtgctgc acgcatggct 2280
gttggctgtg tcatcgagct ggcttccaaa gtggcctcag gagagctgaa gaatgggttt 2340
gctgttgtga ggccccctgg ccatcacgct gaagaatcca cagccatggg gttctgcttt 2400
tttaattcag ttgcaattac cgccaaatac ttgagagacc aactaaatat aagcaagata 2460
ttgattgtag atctggatgt tcaccatgga aacggtaccc agcaggcctt ttatgctgac 2520
cccagcatcc tgtacatttc actccatcgc tatgatgaag ggaacttttt ccctggcagt 2580
ggagccccaa atgaggttcg gtttatttct ttagagcccc acttttattt gtatctttca 2640
ggtaattgca ttgcatgatt accccctaatt ttcttgtcct ttgctggtgt tttaaattac 2700
acgagattac tgaattgtcc catgggacca agaaccagtg cagaacaagt gcataaccca 2760
gagcactgtt tgtcagggaa ggttgggctg atttgatgtg ttgtttgatg tttatttcaa 2820
gagctcccat gtgcttgttt tcctctcttc ttgctttctt ccatttgctc tcttctctgc 2880
ccaccgtggt gtgtctttct cttcccaggt tggaacaggc cttggagaag ggtacaatat 2940
aaatattgcc tggacaggtg gccttgatcc tcccatggga gatgttgagt accttgaagc 3000
attcaggacc atcgtgaagc ctgtggccaa agagtttgat ccagacatgg tcttagtatc 3060
tgctggattt gatgcattgg aaggccacac ccctcctcta ggagggtaca aagtgacggc 3120
aaaatgtttt ggtcatttga cgaagcaatt gatgacattg gctgatggac gtgtggtgtt 3180
ggctctagaa ggaggacatg atctcacagc catctgtgat gcatcagaag cctgtgtaaa 3240
tgcccttcta ggaaatgagc tggagccact tgcagaagat attctccacc aaagcccgaa 3300
tatgaatgct gttatttctt tacagaagat cattgaaatt caaagtatgt ctttaaagtt 3360
ctcttaa                                                            3367
```

<210> 8
<211> 835
<212> PRT
<213> Homo sapiens

<400> 8

```
Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
 1               5                  10                  15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
            20                  25                  30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
        35                  40                  45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
    50                  55                  60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65                  70                  75                  80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
                85                  90                  95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
            100                 105                 110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
        115                 120                 125
```

```
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
    130                 135                 140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145                 150                 155                 160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
                165                 170                 175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
            180                 185                 190
Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
        195                 200                 205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Glu Ser Ser Val Ser Ser
    210                 215                 220
Ser Ser Pro Gly Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly
225                 230                 235                 240
Ser Val Thr Glu Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala
                245                 250                 255
Glu Gln Met Val Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met
            260                 265                 270
Asn Leu Leu Ser Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu
        275                 280                 285
Gly Leu Pro Ala Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys
    290                 295                 300
Glu Lys Gln Lys Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu
305                 310                 315                 320
Pro Gly Gln Tyr Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His
                325                 330                 335
Val Thr Leu Glu Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu
            340                 345                 350
Gln His Leu Leu Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val
        355                 360                 365
Ala Gly Gly Val Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu
    370                 375                 380
Arg Ile Ser Pro Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg
385                 390                 395                 400
Pro Leu Asn Arg Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala
                405                 410                 415
Gln Leu Val Ile Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys
            420                 425                 430
Gln Tyr Gln Gln Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile
        435                 440                 445
Glu Gln Leu Lys Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu
    450                 455                 460
Leu Gln Gly Asp Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly
465                 470                 475                 480
Asn Ser Thr Arg Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly
                485                 490                 495
Gln Val Gly Ala Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu
            500                 505                 510
Asp Ala Gln Ile Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met
        515                 520                 525
Gln Gln Pro Phe Leu Glu Pro Thr His Thr Arg Ala Leu Ser Val Arg
    530                 535                 540
Gln Ala Pro Leu Ala Ala Val Gly Met Asp Gly Leu Glu Lys His Arg
545                 550                 555                 560
Leu Val Ser Arg Thr His Ser Ser Pro Ala Ala Ser Val Leu Pro His
                565                 570                 575
Pro Ala Met Asp Arg Pro Leu Gln Pro Gly Ser Ala Thr Gly Ile Ala
            580                 585                 590
Tyr Asp Pro Leu Met Leu Lys His Gln Cys Val Cys Gly Asn Ser Thr
        595                 600                 605
Thr His Pro Glu His Ala Gly Arg Ile Gln Ser Ile Trp Ser Arg Leu
    610                 615                 620
```

```
Gln Glu Thr Gly Leu Leu Asn Lys Cys Glu Arg Ile Gln Gly Arg Lys
625                 630                 635                 640
Ala Ser Leu Glu Glu Ile Gln Leu Val His Ser Glu His His Ser Leu
                645                 650                 655
Leu Tyr Gly Thr Asn Pro Leu Asp Gly Gln Lys Leu Asp Pro Arg Ile
            660                 665                 670
Leu Leu Gly Asp Asp Ser Gln Lys Phe Phe Ser Ser Leu Pro Cys Gly
        675                 680                 685
Gly Leu Gly Val Asp Ser Asp Thr Ile Trp Asn Glu Leu His Ser Ser
    690                 695                 700
Gly Ala Ala Arg Met Ala Val Gly Cys Val Ile Glu Leu Ala Ser Lys
705                 710                 715                 720
Val Ala Ser Gly Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro
                725                 730                 735
Gly His His Ala Glu Glu Ser Thr Ala Met Gly Phe Cys Phe Phe Asn
            740                 745                 750
Ser Val Ala Ile Thr Ala Lys Tyr Leu Arg Asp Gln Leu Asn Ile Ser
        755                 760                 765
Lys Ile Leu Ile Val Asp Leu Asp Val His His Gly Asn Gly Thr Gln
    770                 775                 780
Gln Ala Phe Tyr Ala Asp Pro Ser Ile Leu Tyr Ile Ser Leu His Arg
785                 790                 795                 800
Tyr Asp Glu Gly Asn Phe Phe Pro Gly Ser Gly Ala Pro Asn Glu Val
                805                 810                 815
Arg Phe Ile Ser Leu Glu Pro His Phe Tyr Leu Tyr Leu Ser Gly Asn
            820                 825                 830
Cys Ile Ala
        835


<210> 9
<211> 1791
<212> DNA
<213> Homo sapiens

<400> 9
ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60
tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tggggtggct 120
ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtggatgtg 180
aagtcagaag ttcctgtggg cctggagccc atctcacctt tagacctaag gacagacctc 240
aggatgatga tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300
cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360
cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420
ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480
caagaacagg aagtagagag gcatcgcaga gaacagcagc ttcctcctct cagaggcaaa 540
gatagaggac gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600
ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660
cgccatccca agctctggta cacggctgcc caccacacat cattggatca aagctctcca 720
ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaag 780
gatgatttcc cccttcgaaa aactgaatcc tcagtcagta gcagttctcc aggctctggt 840
cccagttcac caaacaatgg gccaactgga agtgttactg aaaatgagac ttcggttttg 900
ccccctaccc ctcatgccga gcaaatggtt tcacagcaac gcattctaat tcatgaagat 960
tccatgaacc tgctaagtct ttatacctct ccttctttgc ccaacattac cttggggctt 1020
cccgcagtgc catcccagct caatgcttcg aattcactca aagaaaagca gaagtgtgag 1080
acgcagacgc ttaggcaagg tgttcctctg cctgggcagt atggaggcag catcccggca 1140
tcttccagcc accctcatgt tactttagag ggaaagccac ccaacagcag ccaccaggct 1200
ctcctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct tgtagctggt 1260
ggagttccct tacatcctca gtctcccttg gcaacaaaag agagaatttc acctggcatt 1320
agaggtaccc acaaattgcc ccgtcacaga cccctgaacc gaacccagtc tgcacctttg 1380
cctcagagca cgttggctca gctggtcatt caacagcaac accagcaatt cttggagaag 1440
cagaagcaat accagcagca gatccacatg aacaaactgc tttcgaaatc tattgaacaa 1500
ctgaagcaac caggcagtca ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg 1560
```

atgcaggaag acagagcgcc ctctagtggc aacagcacta ggagcgacag cagtgcttgt 1620
gtggatgaca cactgggaca agttggggct gtgaaggtca aggaggaacc agtggacagt 1680
gatgaagatg ctcagatcca ggaaatggaa tctggggagc aggctgcttt tatgcaacag 1740
gtaataggca aagatttagc tccaggattt gtaattaaag tcattatctg a 1791

<210> 10
<211> 546
<212> PRT
<213> Homo sapiens

<400> 10
Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
1 5 10 15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
20 25 30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
35 40 45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
50 55 60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65 70 75 80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
85 90 95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
100 105 110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
115 120 125
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
130 135 140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145 150 155 160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
165 170 175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
180 185 190
Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
195 200 205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Glu Ser Ser Val Ser Ser
210 215 220
Ser Ser Pro Gly Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly
225 230 235 240
Ser Val Thr Glu Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala
245 250 255
Glu Gln Met Val Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met
260 265 270
Asn Leu Leu Ser Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu
275 280 285
Gly Leu Pro Ala Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys
290 295 300
Glu Lys Gln Lys Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu
305 310 315 320
Pro Gly Gln Tyr Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His
325 330 335
Val Thr Leu Glu Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu
340 345 350
Gln His Leu Leu Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val
355 360 365
Ala Gly Gly Val Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu
370 375 380
Arg Ile Ser Pro Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg
385 390 395 400
Pro Leu Asn Arg Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala

```
                    405                     410                     415
Gln Leu Val Ile Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys
                420                     425                     430
Gln Tyr Gln Gln Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile
            435                     440                     445
Glu Gln Leu Lys Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu
        450                     455                     460
Leu Gln Gly Asp Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly
465                     470                     475                     480
Asn Ser Thr Arg Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly
                    485                     490                     495
Gln Val Gly Ala Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu
                500                     505                     510
Asp Ala Gln Ile Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met
            515                     520                     525
Gln Gln Val Ile Gly Lys Asp Leu Ala Pro Gly Phe Val Ile Lys Val
        530                     535                     540
Ile Ile
545


<210> 11
<211> 590
<212> PRT
<213> Homo sapiens

<400> 11
Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
1               5                   10                  15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
                20                  25                  30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
            35                  40                  45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
        50                  55                  60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65                  70                  75                  80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
                85                  90                  95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
                100                 105                 110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
            115                 120                 125
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
        130                 135                 140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145                 150                 155                 160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
                165                 170                 175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
                180                 185                 190
Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
            195                 200                 205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Ala Ser Glu Pro Asn Leu
        210                 215                 220
Lys Val Arg Ser Arg Leu Lys Gln Lys Val Ala Glu Arg Arg Ser Ser
225                 230                 235                 240
Pro Leu Leu Arg Arg Lys Asp Gly Asn Val Val Thr Ser Phe Lys Lys
                245                 250                 255
Arg Met Phe Glu Val Thr Glu Ser Ser Val Ser Ser Ser Ser Pro Gly
                260                 265                 270
Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly Ser Val Thr Glu
```

```
        275                 280                 285
Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala Glu Gln Met Val
    290                 295                 300
Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met Asn Leu Leu Ser
305                 310                 315                 320
Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu Gly Leu Pro Ala
                325                 330                 335
Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys Glu Lys Gln Lys
            340                 345                 350
Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu Pro Gly Gln Tyr
        355                 360                 365
Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His Val Thr Leu Glu
    370                 375                 380
Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu Gln His Leu Leu
385                 390                 395                 400
Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val Ala Gly Gly Val
                405                 410                 415
Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu Arg Ile Ser Pro
            420                 425                 430
Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg Pro Leu Asn Arg
        435                 440                 445
Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala Gln Leu Val Ile
    450                 455                 460
Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys Gln Tyr Gln Gln
465                 470                 475                 480
Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile Glu Gln Leu Lys
                485                 490                 495
Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu Leu Gln Gly Asp
            500                 505                 510
Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly Asn Ser Thr Arg
        515                 520                 525
Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly Gln Val Gly Ala
    530                 535                 540
Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu Asp Ala Gln Ile
545                 550                 555                 560
Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met Gln Gln Val Ile
                565                 570                 575
Gly Lys Asp Leu Ala Pro Gly Phe Val Ile Lys Val Ile Ile
            580                 585                 590
```

<210> 12
<211> 1084
<212> PRT
<213> Homo sapiens

<400> 12

```
Met Ser Ser Gln Ser His Pro Asp Gly Leu Ser Gly Arg Asp Gln Pro
 1               5                  10                  15
Val Glu Leu Leu Asn Pro Ala Arg Val Asn His Met Pro Ser Thr Val
            20                  25                  30
Asp Val Ala Thr Ala Leu Pro Leu Gln Val Ala Pro Ser Ala Val Pro
        35                  40                  45
Met Asp Leu Arg Leu Asp His Gln Phe Ser Leu Pro Val Ala Glu Pro
    50                  55                  60
Ala Leu Arg Glu Gln Gln Leu Gln Gln Glu Leu Leu Ala Leu Lys Gln
65                  70                  75                  80
Lys Gln Gln Ile Gln Arg Gln Ile Leu Ile Ala Glu Phe Gln Arg Gln
                85                  90                  95
His Glu Gln Leu Ser Arg Gln His Glu Ala Gln Leu His Glu His Ile
            100                 105                 110
Lys Gln Gln Gln Glu Met Leu Ala Met Lys His Gln Gln Glu Leu Leu
```

115 120 125
Glu His Gln Arg Lys Leu Glu Arg His Arg Gln Glu Gln Glu Leu Glu
130 135 140
Lys Gln His Arg Glu Gln Lys Leu Gln Gln Leu Lys Asn Lys Glu Lys
145 150 155 160
Gly Lys Glu Ser Ala Val Ala Ser Thr Glu Val Lys Met Lys Leu Gln
165 170 175
Glu Phe Val Leu Asn Lys Lys Lys Ala Leu Ala His Arg Asn Leu Asn
180 185 190
His Cys Ile Ser Ser Asp Pro Arg Tyr Trp Tyr Gly Lys Thr Gln His
195 200 205
Ser Ser Leu Asp Gln Ser Ser Pro Pro Gln Ser Gly Val Ser Thr Ser
210 215 220
Tyr Asn His Pro Val Leu Gly Met Tyr Asp Ala Lys Asp Asp Phe Pro
225 230 235 240
Leu Arg Lys Thr Ala Ser Glu Pro Asn Leu Lys Leu Arg Ser Arg Leu
245 250 255
Lys Gln Lys Val Ala Glu Arg Arg Ser Ser Pro Leu Leu Arg Arg Lys
260 265 270
Asp Gly Pro Val Val Thr Ala Leu Lys Lys Arg Pro Leu Asp Val Thr
275 280 285
Asp Ser Ala Cys Ser Ser Ala Pro Gly Ser Gly Pro Ser Ser Pro Asn
290 295 300
Asn Ser Ser Gly Ser Val Ser Ala Glu Asn Gly Ile Ala Pro Ala Val
305 310 315 320
Pro Ser Ile Pro Ala Glu Thr Ser Leu Ala His Arg Leu Val Ala Arg
325 330 335
Glu Gly Ser Ala Ala Pro Leu Pro Leu Tyr Thr Ser Pro Ser Leu Pro
340 345 350
Asn Ile Thr Leu Gly Leu Pro Ala Thr Gly Pro Ser Ala Gly Thr Ala
355 360 365
Gly Gln Gln Asp Thr Glu Arg Leu Thr Leu Pro Ala Leu Gln Gln Arg
370 375 380
Leu Ser Leu Phe Pro Gly Thr His Leu Thr Pro Tyr Leu Ser Thr Ser
385 390 395 400
Pro Leu Glu Arg Asp Gly Gly Ala Ala His Ser Pro Leu Leu Gln His
405 410 415
Met Val Leu Leu Glu Gln Pro Pro Ala Gln Ala Pro Leu Val Thr Gly
420 425 430
Leu Gly Ala Leu Pro Leu His Ala Gln Ser Leu Val Gly Ala Asp Arg
435 440 445
Val Ser Pro Ser Ile His Lys Leu Arg Gln His Arg Pro Leu Gly Arg
450 455 460
Thr Gln Ser Ala Pro Leu Pro Gln Asn Ala Gln Ala Leu Gln His Leu
465 470 475 480
Val Ile Gln Gln Gln His Gln Gln Phe Leu Glu Lys His Lys Gln Gln
485 490 495
Phe Gln Gln Gln Gln Leu Gln Met Asn Lys Ile Ile Pro Lys Pro Ser
500 505 510
Glu Pro Ala Arg Gln Pro Glu Ser His Pro Glu Glu Thr Glu Glu Glu
515 520 525
Leu Arg Glu His Gln Ala Leu Leu Asp Glu Pro Tyr Leu Asp Arg Leu
530 535 540
Pro Gly Gln Lys Glu Ala His Ala Gln Ala Gly Val Gln Val Lys Gln
545 550 555 560
Glu Pro Ile Glu Ser Asp Glu Glu Glu Ala Glu Pro Pro Arg Glu Val
565 570 575
Glu Pro Gly Gln Arg Gln Pro Ser Glu Gln Glu Leu Leu Phe Arg Gln
580 585 590
Gln Ala Leu Leu Leu Glu Gln Gln Arg Ile His Gln Leu Arg Asn Tyr
595 600 605
Gln Ala Ser Met Glu Ala Ala Gly Ile Pro Val Ser Phe Gly Gly His

610 615 620
Arg Pro Leu Ser Arg Ala Gln Ser Ser Pro Ala Ser Ala Thr Phe Pro
625 630 635 640
Val Ser Val Gln Glu Pro Pro Thr Lys Pro Arg Phe Thr Thr Gly Leu
645 650 655
Val Tyr Asp Thr Leu Met Leu Lys His Gln Cys Thr Cys Gly Ser Ser
660 665 670
Ser Ser His Pro Glu His Ala Gly Arg Ile Gln Ser Ile Trp Ser Arg
675 680 685
Leu Gln Glu Thr Gly Leu Arg Gly Lys Cys Glu Cys Ile Arg Gly Arg
690 695 700
Lys Ala Thr Leu Glu Glu Leu Gln Thr Val His Ser Glu Ala His Thr
705 710 715 720
Leu Leu Tyr Gly Thr Asn Pro Leu Asn Arg Gln Lys Leu Asp Ser Lys
725 730 735
Lys Leu Leu Gly Ser Leu Ala Ser Val Phe Val Arg Leu Pro Cys Gly
740 745 750
Gly Val Gly Val Asp Ser Asp Thr Ile Trp Asn Glu Val His Ser Ala
755 760 765
Gly Ala Ala Arg Leu Ala Val Gly Cys Val Val Glu Leu Val Phe Lys
770 775 780
Val Ala Thr Gly Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro
785 790 795 800
Gly His His Ala Glu Glu Ser Thr Pro Met Gly Phe Cys Tyr Phe Asn
805 810 815
Ser Val Ala Val Ala Ala Lys Leu Leu Gln Gln Arg Leu Ser Val Ser
820 825 830
Lys Ile Leu Ile Val Asp Trp Asp Val His His Gly Asn Gly Thr Gln
835 840 845
Gln Ala Phe Tyr Ser Asp Pro Ser Val Leu Tyr Met Ser Leu His Arg
850 855 860
Tyr Asp Asp Gly Asn Phe Phe Pro Gly Ser Gly Ala Pro Asp Glu Val
865 870 875 880
Gly Thr Gly Pro Gly Val Gly Phe Asn Val Asn Met Ala Phe Thr Gly
885 890 895
Gly Leu Asp Pro Pro Met Gly Asp Ala Glu Tyr Leu Ala Ala Phe Arg
900 905 910
Thr Val Val Met Pro Ile Ala Ser Glu Phe Ala Pro Asp Val Val Leu
915 920 925
Val Ser Ser Gly Phe Asp Ala Val Glu Gly His Pro Thr Pro Leu Gly
930 935 940
Gly Tyr Asn Leu Ser Ala Arg Cys Phe Gly Tyr Leu Thr Lys Gln Leu
945 950 955 960
Met Gly Leu Ala Gly Gly Arg Ile Val Leu Ala Leu Glu Gly Gly His
965 970 975
Asp Leu Thr Ala Ile Cys Asp Ala Ser Glu Ala Cys Val Ser Ala Leu
980 985 990
Leu Gly Asn Glu Leu Asp Pro Leu Pro Glu Lys Val Leu Gln Gln Arg
995 1000 1005
Pro Asn Ala Asn Ala Val Arg Ser Met Glu Lys Val Met Glu Ile His
1010 1015 1020
Ser Lys Tyr Trp Arg Cys Leu Gln Arg Thr Thr Ser Thr Ala Gly Arg
1025 1030 1035 1040
Ser Leu Ile Glu Ala Gln Thr Cys Glu Asn Glu Glu Ala Glu Thr Val
1045 1050 1055
Thr Ala Met Ala Ser Leu Ser Val Gly Val Lys Pro Ala Glu Lys Arg
1060 1065 1070
Pro Asp Glu Glu Pro Met Glu Glu Glu Pro Pro Leu
1075 1080

<210> 13

<211> 3550
<212> DNA
<213> Homo sapiens

<400> 13
ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60
tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tggggtggct 120
ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtggatgtg 180
aagtcagaag ttcctgtggg cctggagccc atctcacctt tagacctaag gacagacctc 240
aggatgatga tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300
cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360
cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420
ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480
caagaacagg aagtagagag gcatcgcaga gaacagcagc ttcctcctct cagaggcaaa 540
gatagaggac gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600
ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660
cgccatccca agctctggta cacggctgcc caccacacat cattggatca aagctctcca 720
ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaag 780
gatgatttcc cccttcgaaa aactgcctct gagcccaact tgaaggtgcg gtccaggtta 840
aaacagaaag tggcagagag gagaagcagc cccttactca ggcggaagga tggaaatgtt 900
gtcacttcat tcaagaagcg aatgtttgag gtgacagaat cctcagtcag tagcagttct 960
ccaggctctg gtcccagttc accaaacaat gggccaactg gaagtgttac tgaaaatgag 1020
acttcggttt tgccccctac ccctcatgcc gagcaaatgg tttcacagca acgcattcta 1080
attcatgaag attccatgaa cctgctaagt ctttatacct ctccttcttt gcccaacatt 1140
accttggggc ttcccgcagt gccatcccag ctcaatgctt cgaattcact caaagaaaag 1200
cagaagtgtg agacgcagac gcttaggcaa ggtgttcctc tgcctgggca gtatggaggc 1260
agcatcccgg catcttccag ccaccctcat gttactttag agggaaagcc acccaacagc 1320
agccaccagg ctctcctgca gcatttatta ttgaaagaac aaatgcgaca gcaaaagctt 1380
cttgtagctg gtggagttcc cttacatcct cagtctccct tggcaacaaa agagagaatt 1440
tcacctggca ttagaggtac ccacaaattg ccccgtcaca gacccctgaa ccgaacccag 1500
tctgcacctt tgcctcagag cacgttggct cagctggtca ttcaacagca acaccagcaa 1560
ttcttggaga agcagaagca ataccagcag cagatccaca tgaacaaact gctttcgaaa 1620
tctattgaac aactgaagca accaggcagt caccttgagg aagcagagga agagcttcag 1680
ggggaccagg cgatgcagga agacagagcg ccctctagtg gcaacagcac taggagcgac 1740
agcagtgctt gtgtggatga cacactggga caagttgggg ctgtgaaggt caaggaggaa 1800
ccagtggaca gtgatgaaga tgctcagatc caggaaatgg aatctgggga gcaggctgct 1860
tttatgcaac aggtaatagg caaagattta gctccaggat ttgtaattaa agtcattatc 1920
tgacctttcc tggaacccac gcacacacgt gcgctctctg tgcgccaagc tccgctggct 1980
gcggttggca tggatggatt agagaaacac cgtctcgtct ccaggactca ctcttcccct 2040
gctgcctctg ttttacctca cccagcaatg gaccgccccc tccagcctgg ctctgcaact 2100
ggaattgcct atgacccctt gatgctgaaa caccagtgcg tttgtggcaa ttccaccacc 2160
caccctgagc atgctggacg aatacagagt atctggtcac gactgcaaga aactgggctg 2220
ctaaataaat gtgagcgaat tcaaggtcga aaagccagcc tggaggaaat acagcttgtt 2280
cattctgaac atcactcact gttgtatggc accaaccccc tggacggaca gaagctggac 2340
cccaggatac tcctaggtga tgactctcaa aagttttttt cctcattacc ttgtggtgga 2400
cttggggtgg acagtgacac catttggaat gagctacact cgtccggtgc tgcacgcatg 2460
gctgttggct gtgtcatcga gctggcttcc aaagtggcct caggagagct gaagaatggg 2520
tttgctgttg tgaggccccc tggccatcac gctgaagaat ccacagccat ggggttctgc 2580
ttttttaatt cagttgcaat taccgccaaa tacttgagag accaactaaa tataagcaag 2640
atattgattg tagatctgga tgttcaccat ggaaacggta cccagcaggc cttttatgct 2700
gaccccagca tcctgtacat ttcactccat cgctatgatg aagggaactt tttccctggc 2760
agtggagccc caaatgaggt tcggtttatt tctttagagc cccactttta tttgtatctt 2820
tcaggtaatt gcattgcatg attaccccta attttcttgt cctttgctgg tgttttaaat 2880
tacacgagat tactgaattg tcccatggga ccaagaacca gtgcagaaca agtgcataac 2940
ccagagcact gtttgtcagg gaaggttggg ctgatttgat gtgttgtttg atgtttattt 3000
caagagctcc catgtgcttg ttttcctctc ttcttgcttt cttccatttg ctctcttctc 3060
tgcccaccgt ggtgtgtctt tctcttccca ggttggaaca ggccttggag aagggtacaa 3120
tataaatatt gcctggacag gtggccttga tcctcccatg ggagatgttg agtaccttga 3180
agcattcagg accatcgtga agcctgtggc caaagagttt gatccagaca tggtcttagt 3240
atctgctgga tttgatgcat tggaaggcca cacccctcct ctaggagggt acaaagtgac 3300
ggcaaaatgt tttggtcatt tgacgaagca attgatgaca ttggctgatg gacgtgtggt 3360
gttggctcta gaaggaggac atgatctcac agccatctgt gatgcatcag aagcctgtgt 3420

aaatgccctt ctaggaaatg agctggagcc acttgcagaa gatattctcc accaaagccc 3480
gaatatgaat gctgttattt ctttacagaa gatcattgaa attcaaagta tgtctttaaa 3540
gttctcttaa 3550

<210> 14
<211> 7699
<212> DNA
<213> Homo sapiens

<400> 14
cccattcgcc attcaggctg cgcaactgtt gggaagggcg atcggtgcgg gcctcttcgc 60
tattacgcca gctggcgaaa gggggatgtg ctgcaaggcg attaagttgg gtaacgccca 120
gggttttccc agtcacgacg ttgtaaaacg acggccagtg ccaagctgat ctaatcaata 180
ttggccatta gccatattat tcattggtta tatagcataa atcaatattg gctattggcc 240
attgcatacg ttgtatccat atcataatat gtacatttat attggctcat gtccaacatt 300
accgccatgt tgacattgat tattgactag ttattaatag taatcaatta cggggtcatt 360
agttcatagc ccatatatgg agttccgcgt tacataactt acggtaaatg gcccgcctgg 420
cgaccgccca gcgacccccg cccgttgacg tcaatagtga cgtatgttcc catagtaacg 480
ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg 540
gcagtacatc aagtgtatca tatgccaagt ccgcccccta ttgacgtcaa tgacggtaaa 600
tggcccgcct agcattatgc ccagtacatg accttacggg agtttcctac ttggcagtac 660
atctacgtat tagtcatcgc tattaccatg gtgatgcggt tttggcagta caccaatggg 720
cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga cgtcaatggg 780
agtttgtttt ggcaccaaaa tcaacgggac tttccaaaat gtcgtaataa ccccgccccg 840
ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag agctcgttta 900
gtgaaccgtc agaattcaag cttgcggccg cagatctatc gatctgcagg atatcaccat 960
gcacagtatg atcagctcag tggatgtgaa gtcagaagtt cctgtgggcc tggagcccat 1020
ctcaccttta gacctaagga cagacctcag gatgatgatg cccgtggtgg accctgttgt 1080
ccgtgagaag caattgcagc aggaattact tcttatccag cagcagcaac aaatccagaa 1140
gcagcttctg atagcagagt ttcagaaaca gcatgagaac ttgacacggc agcaccaggc 1200
tcagcttcag gagcatatca aggaacttct agccataaaa cagcaacaag aactcctaga 1260
aaaggagcag aaactggagc agcagaggca agaacaggaa gtagagaggc atcgcagaga 1320
acagcagctt cctcctctca gaggcaaaga tagaggacga gaaagggcag tggcaagtac 1380
agaagtaaag cagaagcttc aagagttcct actgagtaaa tcagcaacga aagacactcc 1440
aactaatgga aaaaatcatt ccgtgagccg ccatcccaag ctctggtaca cggctgccca 1500
ccacacatca ttggatcaaa gctctccacc ccttagtgga acatctccat cctacaagta 1560
cacattacca ggagcacaag atgcaaagga tgatttcccc cttcgaaaaa ctgcctctga 1620
gcccaacttg aaggtgcggt ccaggttaaa acagaaagtg gcagagagga gaagcagccc 1680
cttactcagg cggaaggatg gaaatgttgt cacttcattc aagaagcgaa tgtttgaggt 1740
gacagaatcc tcagtcagta gcagttctcc aggctctggt cccagttcac caaacaatgg 1800
gccaactgga agtgttactg aaaatgagac ttcggttttg ccccctaccc ctcatgccga 1860
gcaaatggtt tcacagcaac gcattctaat tcatgaagat tccatgaacc tgctaagtct 1920
ttatacctct ccttctttgc ccaacattac cttggggctt cccgcagtgc catcccagct 1980
caatgcttcg aattcactca aagaaaagca gaagtgtgag acgcagacgc ttaggcaagg 2040
tgttcctctg cctgggcagt atggaggcag catcccggca tcttccagcc accctcatgt 2100
tactttagag ggaaagccac ccaacagcag ccaccaggct ctcctgcagc atttattatt 2160
gaaagaacaa atgcgacagc aaaagcttct tgtagctggt ggagttccct tacatcctca 2220
gtctcccttg gcaacaaaag agagaatttc acctggcatt agaggtaccc acaaattgcc 2280
ccgtcacaga cccctgaacc gaacccagtc tgcacctttg cctcagagca cgttggctca 2340
gctggtcatt caacagcaac accagcaatt cttggagaag cagaagcaat accagcagca 2400
gatccacatg aacaaactgc tttcgaaatc tattgaacaa ctgaagcaac caggcagtca 2460
ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg atgcaggaag acagagcgcc 2520
ctctagtggc aacagcacta ggagcgacag cagtgcttgt gtggatgaca cactgggaca 2580
agttggggct gtgaaggtca aggaggaacc agtggacagt gatgaagatg ctcagatcca 2640
ggaaatggaa tctggggagc aggctgcttt tatgcaacag cctttcctgg aacccacgca 2700
cacacgtgcg ctctctgtgc gccaagctcc gctggctgcg gttggcatgg atggattaga 2760
gaaacaccgt ctcgtctcca ggactcactc ttcccctgct gcctctgttt tacctcaccc 2820
agcaatggac cgcccccctcc agcctggctc tgcaactgga attgcctatg accccttgat 2880
gctgaaacac cagtgcgttt gtggcaattc caccacccac cctgagcatg ctggacgaat 2940
acagagtatc tggtcacgac tgcaagaaac tgggctgcta aataaatgtg agcgaattca 3000
aggtcgaaaa gccagcctgg aggaaataca gcttgttcat tctgaacatc actcactgtt 3060
gtatggcacc aaccccctgg acggacagaa gctggacccc aggatactcc taggtgatga 3120

ctctcaaaag tttttttcct cattaccttg tggtggactt ggggtggaca gtgacaccat 3180
ttggaatgag ctacactcgt ccggtgctgc acgcatggct gttggctgtg tcatcgagct 3240
ggcttccaaa gtggcctcag gagagctgaa gaatgggttt gctgttgtga ggcccccctgg 3300
ccatcacgct gaagaatcca cagccatggg gttctgcttt tttaattcag ttgcaattac 3360
cgccaaatac ttgagagacc aactaaatat aagcaagata ttgattgtag atctggatgt 3420
tcaccatgga aacggtaccc agcaggcctt ttatgctgac cccagcatcc tgtacatttc 3480
actccatcgc tatgatgaag ggaacttttt ccctggcagt ggagccccaa atgaggttgg 3540
aacaggcctt ggagaagggt acaatataaa tattgcctgg acaggtggcc ttgatcctcc 3600
catgggagat gttgagtacc ttgaagcatt caggaccatc gtgaagcctg tggccaaaga 3660
gtttgatcca gacatggtct tagtatctgc tggatttgat gcattggaag gccacacccc 3720
tcctctagga gggtacaaag tgacggcaaa atgttttggt catttgacga agcaattgat 3780
gacattggct gatggacgtg tggtgttggc tctagaagga ggacatgatc tcacagccat 3840
ctgtgatgca tcagaagcct gtgtaaatgc ccttctagga aatgagctgg agccacttgc 3900
agaagatatt ctccaccaaa gcccgaatat gaatgctgtt atttctttac agaagatcat 3960
tgaaattcaa agtatgtctt taaagttctc tggatccggt accagattac aaggacgacg 4020
atgacaagta gatcccgggt ggcatccctg tgacccctcc ccagtgcctc tcctggcctt 4080
ggaagttgcc actccagtgc ccaccagcct tgtcctaata aaattaagtt gcatcatttt 4140
gtctgactag gtgtcctcta taatattatg gggtggaggg gggtggtatg gagcaagggg 4200
cccaagttgg gaagacaacc tgtagggcct gcggggtcta ttcgggaacc aagctggagt 4260
gcagtggcac aatcttggct cactgcaatc tccgcctcct gggttcaagc gattctcctg 4320
cctcagcctc ccgagttgtt gggattccag gcatgcatga ccaggctcag ctaatttttg 4380
tttttttggt agagacgggg tttcaccata ttggccaggc tggtctccaa ctcctaatct 4440
caggtgatct acccaccttg gcctcccaaa ttgctgggat tacaggcgtg aaccactgct 4500
cccttccctg tccttctgat ttaaaataa ctataccagc aggaggacgt ccagacacag 4560
cataggctac ctgccatggc ccaaccggtg ggacatttga gttgcttgct tggcactgtc 4620
ctctcatgcg ttgggtccac tcagtagatg cctgttgaat tgggtacgcg gccagcttct 4680
gtggaatgtg tgtcagttag ggtgtggaaa gtccccaggc tccccagcag gcagaagtat 4740
gcaaagcatg catctcaatt agtcagcaac caggtgtgga aagtccccca ggctccccag 4800
caggcagaag tatgcaaagc atgcatctca attagtcagc aaccatagtc ccgcccctaa 4860
ctccgcccat cccgccccta actccgccca gttccgccca ttctccgccc catggctgac 4920
taattttttt tatttatgca gaggccgagg ccgcctcggc ctctgagcta ttccagaagt 4980
agtgaggagg cttttttgga ggcctaggct tttgcaaaaa gctcctcgag gaactgaaaa 5040
accagaaagt taattcccta tagtgagtcg tattaaattc gtaatcatgg tcatagctgt 5100
ttcctgtgtg aaattgttat ccgctcacaa ttccacacaa catacgagcc ggaagcataa 5160
agtgtaaagc ctggggtgcc taatgagtga gctaactcac attaattgcg ttgcgctcac 5220
tgcccgcttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg 5280
cggggagagg cggtttgcgt attgggcgct cttccgcttc ctcgctcact gactcgctgc 5340
gctcggtcgt tcggctgcgg cgagcggtat cagctcactc aaaggcggta atacggttat 5400
ccacagaatc aggggataac gcaggaaaga acatgtgagc aaaaggccag caaaaggcca 5460
ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc cctgacgagc 5520
atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta taaagatacc 5580
aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg 5640
gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcaatgc tcacgctgta 5700
ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg 5760
ttcagcccga ccgctgcgcc ttatccggta actatcgtct tgagtccaac ccggtaagac 5820
acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag 5880
gcggtgctac agagttcttg aagtggtggc ctaactacgg ctacactaga agaacagtat 5940
ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt agctcttgat 6000
ccggcaaaca aaccaccgct ggtagcggtg gtttttttgt ttgcaagcag cagattacgc 6060
gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacggggtct gacgctcagt 6120
ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt atcaaaaagg atcttcacct 6180
agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat gagtaaactt 6240
ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtctatttc 6300
gttcatccat agttgcctga ctccccgtcg tgtagataac tacgatacgg gagggcttac 6360
catctggccc cagtgctgca atgataccgc gagacccacg ctcaccggct ccagatttat 6420
cagcaataaa ccagccagcc ggaagggccg agcgcagaag tggtcctgca actttatccg 6480
cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttaata 6540
gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt gtcacgctcg tcgtttggta 6600
tggcttcatt cagctccggt tcccaacgat caaggcgagt tacatgatcc cccatgttgt 6660
gcaaaaaagc ggttagctcc ttcggtcctc cgatcgttgt cagaagtaag ttggccgcag 6720
tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg ccatccgtaa 6780
gatgcttttc tgtgactggt gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc 6840

gaccgagttg ctcttgcccg gcgtcaatac gggataatac cgcgccacat agcagaactt 6900
taaaagtgct catcattgga aaacgttctt cggggcgaaa actctcaagg atcttaccgc 6960
tgttgagatc cagttcgatg taacccactc gtgcacccaa ctgatcttca gcatctttta 7020
ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa 7080
taagggcgac acggaaatgt tgaatactca tactcttcct ttttcaatat tattgaagca 7140
tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag aaaaataaac 7200
aaataggggt tccgcgcaca tttccccgaa aagtgccacc tgacgcgccc tgtagcggcg 7260
cattaagcgc ggcgggtgtg gtggttacgc gcagcgtgac cgctacactt gccagcgccc 7320
tagcgcccgc tcctttcgct ttcttccctt cctttctcgc cacgttcgcc ggctttcccc 7380
gtcaagctct aaatcggggc atccctttag ggttccgatt tagtgcttta cggcacctcg 7440
accccaaaaa acttgattag ggtgatggtt cacgtagtgg gccatcgccc tgatagacgg 7500
tttttcgccc tttgacgttg gagtccacgt tctttaatag tggactcttg ttccaaactg 7560
gaacaacact caaccctatc tcggtctatt cttttgattt ataagggatt ttgccgattt 7620
cggcctattg gttaaaaaat gagctgattt aacaaaaatt taacgcgaat tttaacaaaa 7680
tattaaacgt ttacaattt 7699

<210> 15
<211> 7303
<212> DNA
<213> Homo sapiens

<400> 15
cccattcgcc attcaggctg cgcaactgtt gggaagggcg atcggtgcgg gcctcttcgc 60
tattacgcca gctggcgaaa gggggatgtg ctgcaaggcg attaagttgg gtaacgccca 120
gggttttccc agtcacgacg ttgtaaaacg acggccagtg ccaagctgat ctaatcaata 180
ttggccatta gccatattat tcattggtta tatagcataa atcaatattg gctattggcc 240
attgcatacg ttgtatccat atcataatat gtacatttat attggctcat gtccaacatt 300
accgccatgt tgacattgat tattgactag ttattaatag taatcaatta cggggtcatt 360
agttcatagc ccatatatgg agttccgcgt tacataactt acggtaaatg gcccgcctgg 420
cgaccgccca gcgacccccg cccgttgacg tcaatagtga cgtatgttcc catagtaacg 480
ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg 540
gcagtacatc aagtgtatca tatgccaagt ccgcccccta ttgacgtcaa tgacggtaaa 600
tggcccgcct agcattatgc ccagtacatg accttacggg agtttcctac ttggcagtac 660
atctacgtat tagtcatcgc tattaccatg gtgatgcggt tttggcagta caccaatggg 720
cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga cgtcaatggg 780
agtttgtttt ggcaccaaaa tcaacgggac tttccaaaat gtcgtaataa ccccgccccg 840
ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag agctcgttta 900
gtgaaccgtc agaattcaag cttgcggccg cagatctatc gatctgcagg atatcaccat 960
gcacagtatg atcagctcag tggatgtgaa gtcagaagtt cctgtgggcc tggagcccat 1020
ctcaccttta gacctaagga cagacctcag gatgatgatg cccgtggtgg accctgttgt 1080
ccgtgagaag caattgcagc aggaattact tcttatccag cagcagcaac aaatccagaa 1140
gcagcttctg atagcagagt ttcagaaaca gcatgagaac ttgacacggc agcaccaggc 1200
tcagcttcag gagcatatca aggaacttct agccataaaa cagcaacaag aactcctaga 1260
aaaggagcag aaactggagc agcagaggca agaacaggaa gtagagaggc atcgcagaga 1320
acagcagctt cctcctctca gaggcaaaga tagaggacga gaaagggcag tggcaagtac 1380
agaagtaaag cagaagcttc aagagttcct actgagtaaa tcagcaacga aagacactcc 1440
aactaatgga aaaaatcatt ccgtgagccg ccatcccaag ctctggtaca cggctgccca 1500
ccacacatca ttggatcaaa gctctccacc ccttagtgga acatctccat cctacaagta 1560
cacattacca ggagcacaag atgcaaagga tgatttcccc cttcgaaaaa ctgcctctga 1620
gcccaacttg aaggtgcggt ccaggttaaa acagaaagtg gcagagagga gaagcagccc 1680
cttactcagg cggaaggatg gaaatgttgt cacttcattc aagaagcgaa tgtttgaggt 1740
gacagaatcc tcagtcagta gcagttctcc aggctctggt cccagttcac caaacaatgg 1800
gccaactgga agtgttactg aaaatgagac ttcggttttg ccccctaccc ctcatgccga 1860
gcaaatggtt tcacagcaac gcattctaat tcatgaagat tccatgaacc tgctaagtct 1920
ttatacctct ccttctttgc ccaacattac cttggggctt cccgcagtgc catcccagct 1980
caatgcttcg aattcactca aagaaaagca gaagtgtgag acgcagacgc ttaggcaagg 2040
tgttcctctg cctgggcagt atggaggcag catcccggca tcttccagcc accctcatgt 2100
tactttagag ggaaagccac ccaacagcag ccaccaggct ctcctgcagc atttattatt 2160
gaaagaacaa atgcgacagc aaaagcttct tgtagctggt ggagttccct tacatcctca 2220
gtctcccttg gcaacaaaag agagaatttc acctggcatt agaggtaccc acaaattgcc 2280

```
ccgtcacaga cccctgaacc gaacccagtc tgcacctttg cctcagagca cgttggctca 2340
gctggtcatt caacagcaac accagcaatt cttggagaag cagaagcaat accagcagca 2400
gatccacatg aacaaactgc tttcgaaatc tattgaacaa ctgaagcaac caggcagtca 2460
ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg atgcaggaag acagagcgcc 2520
ctctagtggc aacagcacta ggagcgacag cagtgcttgt gtggatgaca cactgggaca 2580
agttggggct gtgaaggtca aggaggaacc agtggacagt gatgaagatg ctcagatcca 2640
ggaaatggaa tctggggagc aggctgcttt tatgcaacag cctttcctgg aacccacgca 2700
cacacgtgcg ctctctgtgc gccaagctcc gctggctgcg gttggcatgg atggattaga 2760
gaaacaccgt ctcgtctcca ggactcactc ttcccctgct gcctctgttt tacctcaccc 2820
agcaatggac cgcccccctcc agcctggctc tgcaactgga attgcctatg accccttgat 2880
gctgaaacac cagtgcgttt gtggcaattc caccacccac cctgagcatg ctggacgaat 2940
acagagtatc tggtcacgac tgcaagaaac tgggctgcta aataaatgtg agcgaattca 3000
aggtcgaaaa gccagcctgg aggaaataca gcttgttcat tctgaacatc actcactgtt 3060
gtatggcacc aaccccctgg acggacagaa gctggacccc aggatactcc taggtgatga 3120
ctctcaaaag tttttttcct cattaccttg tggtggactt ggggtggaca gtgacaccat 3180
ttggaatgag ctacactcgt ccggtgctgc acgcatggct gttggctgtg tcatcgagct 3240
ggcttccaaa gtggcctcag gagagctgaa gaatgggttt gctgttgtga ggccccctgg 3300
ccatcacgct gaagaatcca cagccatggg gttctgcttt tttaattcag ttgcaattac 3360
cgccaaatac ttgagagacc aactaaatat aagcaagata ttgattgtag atctggatgt 3420
tcaccatgga aacggtaccc agcaggcctt ttatgctgac cccagcatcc tgtacatttc 3480
actccatcgc tatgatgaag ggaacttttt ccctggcagt ggagccccaa atgaggttcg 3540
gtttatttct ttagagcccc acttttattt gtatctttca ggtaattgca ttgcaggatc 3600
cggtaccaga ttacaaggac gacgatgaca agtagatccc gggtggcatc cctgtgaccc 3660
ctccccagtg cctctcctgg ccttggaagt tgccactcca gtgcccacca gccttgtcct 3720
aataaaatta agttgcatca ttttgtctga ctaggtgtcc tctataatat tatggggtgg 3780
aggggggtgg tatggagcaa ggggcccaag ttgggaagac aacctgtagg gcctgcgggg 3840
tctattcggg aaccaagctg gagtgcagtg gcacaatctt ggctcactgc aatctccgcc 3900
tcctgggttc aagcgattct cctgcctcag cctcccgagt tgttgggatt ccaggcatgc 3960
atgaccaggc tcagctaatt tttgtttttt tggtagagac ggggtttcac catattggcc 4020
aggctggtct ccaactccta atctcaggtg atctacccac cttggcctcc caaattgctg 4080
ggattacagg cgtgaaccac tgctcccttc cctgtccttc tgattttaaa ataactatac 4140
cagcaggagg acgtccagac acagcatagg ctacctgcca tggcccaacc ggtgggacat 4200
ttgagttgct tgcttggcac tgtcctctca tgcgttgggt ccactcagta gatgcctgtt 4260
gaattgggta cgcggccagc ttctgtggaa tgtgtgtcag ttagggtgtg gaaagtcccc 4320
aggctcccca gcaggcagaa gtatgcaaag catgcatctc aattagtcag caaccaggtg 4380
tggaaagtcc ccaggctccc cagcaggcag aagtatgcaa agcatgcat ctcaattagt 4440
cagcaaccat agtcccgccc ctaactccgc ccatcccgcc cctaactccg cccagttccg 4500
cccattctcc gccccatggc tgactaattt tttttattta tgcagaggcc gaggccgcct 4560
cggcctctga gctattccag aagtagtgag gaggcttttt tggaggccta ggcttttgca 4620
aaaagctcct cgaggaactg aaaaaccaga aagttaattc cctatagtga gtcgtattaa 4680
attcgtaatc atggtcatag ctgtttcctg tgtgaaattg ttatccgctc acaattccac 4740
acaacatacg agccggaagc ataaagtgta aagcctgggg tgcctaatga gtgagctaac 4800
tcacattaat tgcgttgcgc tcactgcccg ctttccagtc gggaaacctg tcgtgccagc 4860
tgcattaatg aatcggccaa cgcgcgggga gaggcggttt gcgtattggg cgctcttccg 4920
cttcctcgct cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc 4980
actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt 5040
gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc 5100
ataggctccg cccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa 5160
acccgacagg actataaaga taccaggcgt ttccccctgg aagctccctc gtgcgctctc 5220
ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg 5280
cgctttctca atgctcacgc tgtaggtatc tcagttcggt gtaggtcgtt cgctccaagc 5340
tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc 5400
gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca 5460
ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact 5520
acggctacac tagaagaaca gtatttggta tctgcgctct gctgaagcca gttaccttcg 5580
gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt 5640
ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct 5700
tttctacggg gtctgacgct cagtggaacg aaaactcacg ttaagggatt ttggtcatga 5760
gattatcaaa aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttaaatcaa 5820
tctaaagtat atatgagtaa acttggtctg acagttacca atgcttaatc agtgaggcac 5880
ctatctcagc gatctgtcta tttcgttcat ccatagttgc ctgactcccc gtcgtgtaga 5940
taactacgat acgggagggc ttaccatctg gccccagtgc tgcaatgata ccgcgagacc 6000
```

cacgctcacc ggctccagat ttatcagcaa taaaccagcc agccggaagg gccgagcgca 6060
gaagtggtcc tgcaacttta tccgcctcca tccagtctat taattgttgc cgggaagcta 6120
gagtaagtag ttcgccagtt aatagtttgc gcaacgttgt tgccattgct acaggcatcg 6180
tggtgtcacg ctcgtcgttt ggtatggctt cattcagctc cggttcccaa cgatcaaggc 6240
gagttacatg atcccccatg ttgtgcaaaa aagcggttag ctccttcggt cctccgatcg 6300
ttgtcagaag taagttggcc gcagtgttat cactcatggt tatggcagca ctgcataatt 6360
ctcttactgt catgccatcc gtaagatgct tttctgtgac tggtgagtac tcaaccaagt 6420
cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgtca atacgggata 6480
ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt tcttcggggc 6540
gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaaccc actcgtgcac 6600
ccaactgatc ttcagcatct tttactttca ccagcgtttc tgggtgagca aaaacaggaa 6660
ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa atgttgaata ctcatactct 6720
tcctttttca atattattga agcatttatc agggttattg tctcatgagc ggatacatat 6780
ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc cgaaaagtgc 6840
cacctgacgc gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt acgcgcagcg 6900
tgaccgctac acttgccagc gccctagcgc ccgctccttt cgctttcttc ccttcctttc 6960
tcgccacgtt cgccggcttt ccccgtcaag ctctaaatcg ggcatccct ttagggttcc 7020
gatttagtgc tttacggcac ctcgacccca aaaaacttga ttagggtgat ggttcacgta 7080
gtgggccatc gccctgatag acggtttttc gccctttgac gttggagtcc acgttcttta 7140
atagtggact cttgttccaa actggaacaa cactcaaccc tatctcggtc tattcttttg 7200
atttataagg gattttgccg atttcggcct attggttaaa aaatgagctg atttaacaaa 7260
aatttaacgc gaattttaac aaaatattaa acgtttacaa ttt 7303

<210> 16
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer used to amplify human DNA

<400> 16
ccatggaaac ggtacccagc aggc 24

<210> 17
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer used to amplify human DNA

<400> 17
cactccatcg ctatgatgaa ggg 23

<210> 18
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer used to amplify human DNA

<400> 18
agttcccttc atcatagcga tgg 23

<210> 19
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer used to amplify human DNA

<400> 19
aatgtacagg atgctggggt 20

<210> 20
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer used to amplify human DNA

<400> 20
cccttgtagc tggtggagtt ccctt 25

<210> 21
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer used to amplify human DNA

<400> 21
tgtgtcatcg agctggcttc 20

<210> 22
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer used to amplify human DNA

<400> 22
atcttctgca agtggctcca 20

```
Thr His Ser Ser Pro Ala Ala Ser Val Leu Pro His Pro Ala Met Asp
    610                 615                 620
Arg Pro Leu Gln Pro Gly Ser Ala Thr Gly Ile Ala Tyr Asp Pro Leu
625                 630                 635                 640
Met Leu Lys His Gln Cys Val Cys Gly Asn Ser Thr Thr His Pro Glu
                645                 650                 655
His Ala Gly Arg Ile Gln Ser Ile Trp Ser Arg Leu Gln Glu Thr Gly
            660                 665                 670
Leu Leu Asn Lys Cys Glu Arg Ile Gln Gly Arg Lys Ala Ser Leu Glu
        675                 680                 685
Glu Ile Gln Leu Val His Ser Glu His His Ser Leu Leu Tyr Gly Thr
    690                 695                 700
Asn Pro Leu Asp Gly Gln Lys Leu Asp Pro Arg Ile Leu Leu Gly Asp
705                 710                 715                 720
Asp Ser Gln Lys Phe Phe Ser Ser Leu Pro Cys Gly Gly Leu Gly Val
                725                 730                 735
Asp Ser Asp Thr Ile Trp Asn Glu Leu His Ser Ser Gly Ala Ala Arg
            740                 745                 750
Met Ala Val Gly Cys Val Ile Glu Leu Ala Ser Lys Val Ala Ser Gly
        755                 760                 765
Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro Gly His His Ala
    770                 775                 780
Glu Glu Ser Thr Ala Met Gly Phe Cys Phe Phe Asn Ser Val Ala Ile
785                 790                 795                 800
Thr Ala Lys Tyr Leu Arg Asp Gln Leu Asn Ile Ser Lys Ile Leu Ile
                805                 810                 815
Val Asp Leu Asp Val His His Gly Asn Gly Thr Gln Gln Ala Phe Tyr
            820                 825                 830
Ala Asp Pro Ser Ile Leu Tyr Ile Ser Leu His Arg Tyr Asp Glu Gly
        835                 840                 845
Asn Phe Phe Pro Gly Ser Gly Ala Pro Asn Glu Val Arg Phe Ile Ser
    850                 855                 860
Leu Glu Pro His Phe Tyr Leu Tyr Leu Ser Gly Asn Cys Ile Ala
865                 870                 875
```

<210> 5
<211> 3054
<212> DNA
<213> Homo sapiens

<400> 5

```
ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60
tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tggggtggct 120
ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtggatgtg 180
aagtcagaag ttcctgtggg cctggagccc atctcacctt tagacctaag gacagacctc 240
aggatgatga tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300
cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360
cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420
ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480
caagaacagg aagtagagag gcatcgcaga gaacagcagc ttcctcctct cagaggcaaa 540
gatagaggac gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600
ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660
cgccatccca agctctggta cacggctgcc caccacacat cattggatca aagctctcca 720
ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaag 780
gatgatttcc cccttcgaaa aactgaatcc tcagtcagta gcagttctcc aggctctggt 840
cccagttcac caaacaatgg gccaactgga agtgttactg aaaatgagac ttcggttttg 900
ccccctaccc ctcatgccga gcaaatggtt tcacagcaac gcattctaat tcatgaagat 960
tccatgaacc tgctaagtct ttatacctct ccttctttgc ccaacattac cttggggctt 1020
cccgcagtgc catcccagct caatgcttcg aattcactca aagaaaagca gaagtgtgag 1080
acgcagacgc ttaggcaagg tgttcctctg cctgggcagt atggaggcag catcccggca 1140
tcttccagcc accctcatgt tactttagag ggaaagccac ccaacagcag ccaccaggct 1200
```

```
ctcctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct tgtagctggt 1260
ggagttccct tacatcctca gtctcccttg gcaacaaaag agagaatttc acctggcatt 1320
agaggtaccc acaaattgcc ccgtcacaga cccctgaacc gaacccagtc tgcacctttg 1380
cctcagagca cgttggctca gctggtcatt caacagcaac accagcaatt cttggagaag 1440
cagaagcaat accagcagca gatccacatg aacaaactgc tttcgaaatc tattgaacaa 1500
ctgaagcaac caggcagtca ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg 1560
atgcaggaag acagagcgcc ctctagtggc aacagcacta ggagcgacag cagtgcttgt 1620
gtggatgaca cactgggaca agttggggct gtgaaggtca aggaggaacc agtggacagt 1680
gatgaagatg ctcagatcca ggaaatggaa tctggggagc aggctgcttt tatgcaacag 1740
cctttcctgg aacccacgca cacacgtgcg ctctctgtgc gccaagctcc gctggctgcg 1800
gttggcatgg atggattaga gaaacaccgt ctcgtctcca ggactcactc ttcccctgct 1860
gcctctgttt tacctcaccc agcaatggac cgcccccctcc agcctggctc tgcaactgga 1920
attgcctatg acccccttgat gctgaaacac cagtgcgttt gtggcaattc caccacccac 1980
cctgagcatg ctggacgaat acagagtatc tggtcacgac tgcaagaaac tgggctgcta 2040
aataaatgtg agcgaattca aggtcgaaaa gccagcctgg aggaaataca gcttgttcat 2100
tctgaacatc actcactgtt gtatggcacc aaccccctgg acggacagaa gctggacccc 2160
aggatactcc taggtgatga ctctcaaaag tttttttcct cattaccttg tggtggactt 2220
ggggtggaca gtgacaccat ttggaatgag ctacactcgt ccggtgctgc acgcatggct 2280
gttggctgtg tcatcgagct ggcttccaaa gtggcctcag gagagctgaa gaatgggttt 2340
gctgttgtga ggccccctgg ccatcacgct gaagaatcca cagccatggg gttctgcttt 2400
tttaattcag ttgcaattac cgccaaatac ttgagagacc aactaaatat aagcaagata 2460
ttgattgtag atctggatgt tcaccatgga aacggtaccc agcaggcctt ttatgctgac 2520
cccagcatcc tgtacatttc actccatcgc tatgatgaag ggaacttttt ccctggcagt 2580
ggagccccaa atgaggttgg aacaggcctt ggagaagggt acaatataaa tattgcctgg 2640
acaggtggcc ttgatcctcc catgggagat gttgagtacc ttgaagcatt caggaccatc 2700
gtgaagcctg tggccaaaga gtttgatcca gacatggtct tagtatctgc tggatttgat 2760
gcattggaag gccacacccc tcctctagga gggtacaaag tgacggcaaa atgttttggt 2820
catttgacga agcaattgat gacattggct gatggacgtg tggtgttggc tctagaagga 2880
ggacatgatc tcacagccat ctgtgatgca tcagaagcct gtgtaaatgc ccttctagga 2940
aatgagctgg agccacttgc agaagatatt ctccaccaaa gcccgaatat gaatgctgtt 3000
atttctttac agaagatcat tgaaattcaa agtatgtctt taaagttctc ttaa       3054
```

<210> 6
<211> 967
<212> PRT
<213> Homo sapiens

<400> 6

```
Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
1               5                   10                  15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
            20                  25                  30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
        35                  40                  45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
    50                  55                  60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65                  70                  75                  80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
                85                  90                  95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
            100                 105                 110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
        115                 120                 125
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
    130                 135                 140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145                 150                 155                 160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
                165                 170                 175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
            180                 185                 190
```

Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
195 200 205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Glu Ser Ser Val Ser Ser
210 215 220
Ser Ser Pro Gly Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly
225 230 235 240
Ser Val Thr Glu Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala
245 250 255
Glu Gln Met Val Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met
260 265 270
Asn Leu Leu Ser Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu
275 280 285
Gly Leu Pro Ala Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys
290 295 300
Glu Lys Gln Lys Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu
305 310 315 320
Pro Gly Gln Tyr Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His
325 330 335
Val Thr Leu Glu Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu
340 345 350
Gln His Leu Leu Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val
355 360 365
Ala Gly Gly Val Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu
370 375 380
Arg Ile Ser Pro Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg
385 390 395 400
Pro Leu Asn Arg Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala
405 410 415
Gln Leu Val Ile Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys
420 425 430
Gln Tyr Gln Gln Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile
435 440 445
Glu Gln Leu Lys Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu
450 455 460
Leu Gln Gly Asp Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly
465 470 475 480
Asn Ser Thr Arg Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly
485 490 495
Gln Val Gly Ala Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu
500 505 510
Asp Ala Gln Ile Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met
515 520 525
Gln Gln Pro Phe Leu Glu Pro Thr His Thr Arg Ala Leu Ser Val Arg
530 535 540
Gln Ala Pro Leu Ala Ala Val Gly Met Asp Gly Leu Glu Lys His Arg
545 550 555 560
Leu Val Ser Arg Thr His Ser Ser Pro Ala Ala Ser Val Leu Pro His
565 570 575
Pro Ala Met Asp Arg Pro Leu Gln Pro Gly Ser Ala Thr Gly Ile Ala
580 585 590
Tyr Asp Pro Leu Met Leu Lys His Gln Cys Val Cys Gly Asn Ser Thr
595 600 605
Thr His Pro Glu His Ala Gly Arg Ile Gln Ser Ile Trp Ser Arg Leu
610 615 620
Gln Glu Thr Gly Leu Leu Asn Lys Cys Glu Arg Ile Gln Gly Arg Lys
625 630 635 640
Ala Ser Leu Glu Glu Ile Gln Leu Val His Ser Glu His His Ser Leu
645 650 655
Leu Tyr Gly Thr Asn Pro Leu Asp Gly Gln Lys Leu Asp Pro Arg Ile
660 665 670
Leu Leu Gly Asp Asp Ser Gln Lys Phe Phe Ser Ser Leu Pro Cys Gly
675 680 685

```
Gly Leu Gly Val Asp Ser Asp Thr Ile Trp Asn Glu Leu His Ser Ser
    690                 695                 700
Gly Ala Ala Arg Met Ala Val Gly Cys Val Ile Glu Leu Ala Ser Lys
705                 710                 715                 720
Val Ala Ser Gly Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro
                725                 730                 735
Gly His His Ala Glu Glu Ser Thr Ala Met Gly Phe Cys Phe Phe Asn
            740                 745                 750
Ser Val Ala Ile Thr Ala Lys Tyr Leu Arg Asp Gln Leu Asn Ile Ser
        755                 760                 765
Lys Ile Leu Ile Val Asp Leu Asp Val His His Gly Asn Gly Thr Gln
    770                 775                 780
Gln Ala Phe Tyr Ala Asp Pro Ser Ile Leu Tyr Ile Ser Leu His Arg
785                 790                 795                 800
Tyr Asp Glu Gly Asn Phe Phe Pro Gly Ser Gly Ala Pro Asn Glu Val
                805                 810                 815
Gly Thr Gly Leu Gly Glu Gly Tyr Asn Ile Asn Ile Ala Trp Thr Gly
            820                 825                 830
Gly Leu Asp Pro Pro Met Gly Asp Val Glu Tyr Leu Glu Ala Phe Arg
        835                 840                 845
Thr Ile Val Lys Pro Val Ala Lys Glu Phe Asp Pro Asp Met Val Leu
    850                 855                 860
Val Ser Ala Gly Phe Asp Ala Leu Glu Gly His Thr Pro Pro Leu Gly
865                 870                 875                 880
Gly Tyr Lys Val Thr Ala Lys Cys Phe Gly His Leu Thr Lys Gln Leu
                885                 890                 895
Met Thr Leu Ala Asp Gly Arg Val Val Leu Ala Leu Glu Gly Gly His
            900                 905                 910
Asp Leu Thr Ala Ile Cys Asp Ala Ser Glu Ala Cys Val Asn Ala Leu
        915                 920                 925
Leu Gly Asn Glu Leu Glu Pro Leu Ala Glu Asp Ile Leu His Gln Ser
    930                 935                 940
Pro Asn Met Asn Ala Val Ile Ser Leu Gln Lys Ile Ile Glu Ile Gln
945                 950                 955                 960
Ser Met Ser Leu Lys Phe Ser
                965
```

<210> 7
<211> 3367
<212> DNA
<213> Homo sapiens

<400> 7

```
ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60
tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tggggtggct 120
ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtggatgtg 180
aagtcagaag ttcctgtggg cctggagccc atctcacctt tagacctaag gacagacctc 240
aggatgatga tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300
cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360
cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420
ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480
caagaacagg aagtagagag gcatcgcaga gaacagcagc ttcctcctct cagaggcaaa 540
gatagaggac gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600
ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660
cgccatccca agctctggta cacggctgcc caccacacat cattggatca aagctctcca 720
ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaag 780
gatgatttcc cccttcgaaa aactgaatcc tcagtcagta gcagttctcc aggctctggt 840
cccagttcac caaacaatgg gccaactgga agtgttactg aaaatgagac ttcggttttg 900
ccccctaccc ctcatgccga gcaaatggtt tcacagcaac gcattctaat tcatgaagat 960
tccatgaacc tgctaagtct ttatacctct ccttctttgc ccaacattac cttggggctt 1020
cccgcagtgc catcccagct caatgcttcg aattcactca aagaaaagca gaagtgtgag 1080
```

acgcagacgc ttaggcaagg tgttcctctg cctgggcagt atggaggcag catcccggca 1140
tcttccagcc accctcatgt tactttagag ggaaagccac ccaacagcag ccaccaggct 1200
ctcctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct tgtagctggt 1260
ggagttccct tacatcctca gtctcccttg gcaacaaaag agagaatttc acctggcatt 1320
agaggtaccc acaaattgcc ccgtcacaga cccctgaacc gaacccagtc tgcacctttg 1380
cctcagagca cgttggctca gctggtcatt caacagcaac accagcaatt cttggagaag 1440
cagaagcaat accagcagca gatccacatg aacaaactgc tttcgaaatc tattgaacaa 1500
ctgaagcaac caggcagtca ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg 1560
atgcaggaag acagagcgcc ctctagtggc aacagcacta ggagcgacag cagtgcttgt 1620
gtggatgaca cactgggaca agttggggct gtgaaggtca aggaggaacc agtggacagt 1680
gatgaagatg ctcagatcca ggaaatggaa tctggggagc aggctgcttt tatgcaacag 1740
cctttcctgg aacccacgca cacacgtgcg ctctctgtgc gccaagctcc gctggctgcg 1800
gttggcatgg atggattaga gaaacaccgt ctcgtctcca ggactcactc ttcccctgct 1860
gcctctgttt tacctcaccc agcaatggac cgcccccctcc agcctggctc tgcaactgga 1920
attgcctatg acccccttgat gctgaaacac cagtgcgttt gtggcaattc caccacccac 1980
cctgagcatg ctggacgaat acagagtatc tggtcacgac tgcaagaaac tgggctgcta 2040
aataaatgtg agcgaattca aggtcgaaaa gccagcctgg aggaaataca gcttgttcat 2100
tctgaacatc actcactgtt gtatggcacc aacccccctgg acggacagaa gctggacccc 2160
aggatactcc taggtgatga ctctcaaaag tttttttcct cattaccttg tggtggactt 2220
ggggtggaca gtgacaccat ttggaatgag ctacactcgt ccggtgctgc acgcatggct 2280
gttggctgtg tcatcgagct ggcttccaaa gtggcctcag gagagctgaa gaatgggttt 2340
gctgttgtga ggcccccctgg ccatcacgct gaagaatcca cagccatggg gttctgcttt 2400
tttaattcag ttgcaattac cgccaaatac ttgagagacc aactaaatat aagcaagata 2460
ttgattgtag atctggatgt tcaccatgga aacggtaccc agcaggcctt ttatgctgac 2520
cccagcatcc tgtacatttc actccatcgc tatgatgaag ggaacttttt ccctggcagt 2580
ggagccccaa atgaggttcg gtttatttct ttagagcccc acttttattt gtatctttca 2640
ggtaattgca ttgcatgatt acccctaatt ttcttgtcct ttgctggtgt tttaaattac 2700
acgagattac tgaattgtcc catgggacca agaaccagtg cagaacaagt gcataaccca 2760
gagcactgtt tgtcagggaa ggttgggctg atttgatgtg ttgtttgatg tttatttcaa 2820
gagctcccat gtgcttgttt tcctctcttc ttgctttctt ccatttgctc tcttctctgc 2880
ccaccgtggt gtgtctttct cttcccaggt tggaacaggc cttggagaag ggtacaatat 2940
aaatattgcc tggacaggtg gccttgatcc tcccatggga gatgttgagt accttgaagc 3000
attcaggacc atcgtgaagc ctgtggccaa agagtttgat ccagacatgg tcttagtatc 3060
tgctggattt gatgcattgg aaggccacac ccctcctcta ggagggtaca aagtgacggc 3120
aaaatgtttt ggtcatttga cgaagcaatt gatgacattg gctgatggac gtgtggtgtt 3180
ggctctagaa ggaggacatg atctcacagc catctgtgat gcatcagaag cctgtgtaaa 3240
tgcccttcta ggaaatgagc tggagccact tgcagaagat attctccacc aaagcccgaa 3300
tatgaatgct gttatttctt tacagaagat cattgaaatt caaagtatgt ctttaaagtt 3360
ctcttaa 3367

<210> 8
<211> 835
<212> PRT
<213> Homo sapiens

<400> 8
Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
1 5 10 15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
20 25 30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
35 40 45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
50 55 60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65 70 75 80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
85 90 95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
100 105 110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
115 120 125

```
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
    130                 135                 140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145                 150                 155                 160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
                165                 170                 175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
            180                 185                 190
Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
        195                 200                 205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Glu Ser Ser Val Ser Ser
    210                 215                 220
Ser Ser Pro Gly Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly
225                 230                 235                 240
Ser Val Thr Glu Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala
                245                 250                 255
Glu Gln Met Val Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met
            260                 265                 270
Asn Leu Leu Ser Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu
        275                 280                 285
Gly Leu Pro Ala Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys
    290                 295                 300
Glu Lys Gln Lys Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu
305                 310                 315                 320
Pro Gly Gln Tyr Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His
                325                 330                 335
Val Thr Leu Glu Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu
            340                 345                 350
Gln His Leu Leu Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val
        355                 360                 365
Ala Gly Gly Val Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu
    370                 375                 380
Arg Ile Ser Pro Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg
385                 390                 395                 400
Pro Leu Asn Arg Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala
                405                 410                 415
Gln Leu Val Ile Gln Gln Gln His Gln Phe Leu Glu Lys Gln Lys
            420                 425                 430
Gln Tyr Gln Gln Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile
        435                 440                 445
Glu Gln Leu Lys Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu
    450                 455                 460
Leu Gln Gly Asp Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly
465                 470                 475                 480
Asn Ser Thr Arg Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly
                485                 490                 495
Gln Val Gly Ala Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu
            500                 505                 510
Asp Ala Gln Ile Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met
        515                 520                 525
Gln Gln Pro Phe Leu Glu Pro Thr His Thr Arg Ala Leu Ser Val Arg
    530                 535                 540
Gln Ala Pro Leu Ala Ala Val Gly Met Asp Gly Leu Glu Lys His Arg
545                 550                 555                 560
Leu Val Ser Arg Thr His Ser Ser Pro Ala Ala Ser Val Leu Pro His
                565                 570                 575
Pro Ala Met Asp Arg Pro Leu Gln Pro Gly Ser Ala Thr Gly Ile Ala
            580                 585                 590
Tyr Asp Pro Leu Met Leu Lys His Gln Cys Val Cys Gly Asn Ser Thr
        595                 600                 605
Thr His Pro Glu His Ala Gly Arg Ile Gln Ser Ile Trp Ser Arg Leu
    610                 615                 620
```

```
Gln Glu Thr Gly Leu Leu Asn Lys Cys Glu Arg Ile Gln Gly Arg Lys
625                 630                 635                 640
Ala Ser Leu Glu Glu Ile Gln Leu Val His Ser Glu His His Ser Leu
                645                 650                 655
Leu Tyr Gly Thr Asn Pro Leu Asp Gly Gln Lys Leu Asp Pro Arg Ile
            660                 665                 670
Leu Leu Gly Asp Asp Ser Gln Lys Phe Phe Ser Ser Leu Pro Cys Gly
        675                 680                 685
Gly Leu Gly Val Asp Ser Asp Thr Ile Trp Asn Glu Leu His Ser Ser
    690                 695                 700
Gly Ala Ala Arg Met Ala Val Gly Cys Val Ile Glu Leu Ala Ser Lys
705                 710                 715                 720
Val Ala Ser Gly Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro
                725                 730                 735
Gly His His Ala Glu Glu Ser Thr Ala Met Gly Phe Cys Phe Phe Asn
            740                 745                 750
Ser Val Ala Ile Thr Ala Lys Tyr Leu Arg Asp Gln Leu Asn Ile Ser
        755                 760                 765
Lys Ile Leu Ile Val Asp Leu Asp Val His His Gly Asn Gly Thr Gln
    770                 775                 780
Gln Ala Phe Tyr Ala Asp Pro Ser Ile Leu Tyr Ile Ser Leu His Arg
785                 790                 795                 800
Tyr Asp Glu Gly Asn Phe Phe Pro Gly Ser Gly Ala Pro Asn Glu Val
                805                 810                 815
Arg Phe Ile Ser Leu Glu Pro His Phe Tyr Leu Tyr Leu Ser Gly Asn
            820                 825                 830
Cys Ile Ala
        835
```

<210> 9
<211> 1791
<212> DNA
<213> Homo sapiens

<400> 9

```
ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60
tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tggggtggct 120
ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtggatgtg 180
aagtcagaag ttcctgtggg cctggagccc atctcacctt tagacctaag gacagacctc 240
aggatgatga tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300
cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360
cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420
ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480
caagaacagg aagtagagag gcatcgcaga gaacagcagc ttcctcctct cagaggcaaa 540
gatagaggac gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600
ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660
cgccatccca agctctggta cacggctgcc caccacacat cattggatca aagctctcca 720
ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaag 780
gatgatttcc cccttcgaaa aactgaatcc tcagtcagta gcagttctcc aggctctggt 840
cccagttcac caaacaatgg gccaactgga agtgttactg aaaatgagac ttcggttttg 900
ccccctaccc ctcatgccga gcaaatggtt tcacagcaac gcattctaat tcatgaagat 960
tccatgaacc tgctaagtct ttatacctct ccttctttgc ccaacattac cttggggctt 1020
cccgcagtgc catcccagct caatgcttcg aattcactca aagaaaagca gaagtgtgag 1080
acgcagacgc ttaggcaagg tgttcctctg cctgggcagt atggaggcag catcccggca 1140
tcttccagcc accctcatgt tactttagag ggaaagccac ccaacagcag ccaccaggct 1200
ctcctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct tgtagctggt 1260
ggagttccct tacatcctca gtctcccttg gcaacaaaag agagaatttc acctggcatt 1320
agaggtaccc acaaattgcc ccgtcacaga cccctgaacc gaacccagtc tgcacctttg 1380
cctcagagca cgttggctca gctggtcatt caacagcaac accagcaatt cttggagaag 1440
cagaagcaat accagcagca gatccacatg aacaaactgc tttcgaaatc tattgaacaa 1500
ctgaagcaac caggcagtca ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg 1560
```

```
atgcaggaag acagagcgcc ctctagtggc aacagcacta ggagcgacag cagtgcttgt 1620
gtggatgaca cactgggaca agttggggct gtgaaggtca aggaggaacc agtggacagt 1680
gatgaagatg ctcagatcca ggaaatggaa tctggggagc aggctgcttt tatgcaacag 1740
gtaataggca aagatttagc tccaggattt gtaattaaag tcattatctg a          1791

<210> 10
<211> 546
<212> PRT
<213> Homo sapiens

<400> 10
Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
 1               5                   10                  15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
            20                  25                  30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
        35                  40                  45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
    50                  55                  60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65                  70                  75                  80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
                85                  90                  95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
            100                 105                 110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
        115                 120                 125
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
    130                 135                 140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145                 150                 155                 160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
                165                 170                 175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
            180                 185                 190
Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
        195                 200                 205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Glu Ser Ser Val Ser Ser
    210                 215                 220
Ser Ser Pro Gly Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly
225                 230                 235                 240
Ser Val Thr Glu Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala
                245                 250                 255
Glu Gln Met Val Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met
            260                 265                 270
Asn Leu Leu Ser Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu
        275                 280                 285
Gly Leu Pro Ala Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys
    290                 295                 300
Glu Lys Gln Lys Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu
305                 310                 315                 320
Pro Gly Gln Tyr Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His
                325                 330                 335
Val Thr Leu Glu Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu
            340                 345                 350
Gln His Leu Leu Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val
        355                 360                 365
Ala Gly Gly Val Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu
    370                 375                 380
Arg Ile Ser Pro Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg
385                 390                 395                 400
Pro Leu Asn Arg Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala
```

```
                        405                     410                     415
Gln Leu Val Ile Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys
            420                     425                     430
Gln Tyr Gln Gln Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile
        435                     440                     445
Glu Gln Leu Lys Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu
    450                     455                     460
Leu Gln Gly Asp Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly
465                     470                     475                     480
Asn Ser Thr Arg Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly
                485                     490                     495
Gln Val Gly Ala Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu
            500                     505                     510
Asp Ala Gln Ile Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met
        515                     520                     525
Gln Gln Val Ile Gly Lys Asp Leu Ala Pro Gly Phe Val Ile Lys Val
    530                     535                     540
Ile Ile
545


<210> 11
<211> 590
<212> PRT
<213> Homo sapiens

<400> 11
Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
 1               5                  10                  15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
            20                  25                  30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
        35                  40                  45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
    50                  55                  60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65                  70                  75                  80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
                85                  90                  95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
            100                 105                 110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
        115                 120                 125
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
    130                 135                 140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145                 150                 155                 160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
                165                 170                 175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
            180                 185                 190
Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
        195                 200                 205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Ala Ser Glu Pro Asn Leu
    210                 215                 220
Lys Val Arg Ser Arg Leu Lys Gln Lys Val Ala Glu Arg Arg Ser Ser
225                 230                 235                 240
Pro Leu Leu Arg Arg Lys Asp Gly Asn Val Val Thr Ser Phe Lys Lys
                245                 250                 255
Arg Met Phe Glu Val Thr Glu Ser Ser Val Ser Ser Ser Ser Pro Gly
            260                 265                 270
Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly Ser Val Thr Glu
```

```
        275                 280                 285
Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala Glu Gln Met Val
    290                 295                 300
Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met Asn Leu Leu Ser
305                 310                 315                 320
Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu Gly Leu Pro Ala
                325                 330                 335
Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys Glu Lys Gln Lys
            340                 345                 350
Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu Pro Gly Gln Tyr
        355                 360                 365
Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His Val Thr Leu Glu
    370                 375                 380
Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu Gln His Leu Leu
385                 390                 395                 400
Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val Ala Gly Gly Val
                405                 410                 415
Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu Arg Ile Ser Pro
            420                 425                 430
Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg Pro Leu Asn Arg
        435                 440                 445
Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala Gln Leu Val Ile
    450                 455                 460
Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys Gln Tyr Gln Gln
465                 470                 475                 480
Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile Glu Gln Leu Lys
                485                 490                 495
Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu Leu Gln Gly Asp
            500                 505                 510
Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly Asn Ser Thr Arg
        515                 520                 525
Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly Gln Val Gly Ala
    530                 535                 540
Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu Asp Ala Gln Ile
545                 550                 555                 560
Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met Gln Gln Val Ile
                565                 570                 575
Gly Lys Asp Leu Ala Pro Gly Phe Val Ile Lys Val Ile Ile
            580                 585                 590
```

```
<210> 12
<211> 1084
<212> PRT
<213> Homo sapiens

<400> 12
Met Ser Ser Gln Ser His Pro Asp Gly Leu Ser Gly Arg Asp Gln Pro
 1               5                  10                  15
Val Glu Leu Leu Asn Pro Ala Arg Val Asn His Met Pro Ser Thr Val
            20                  25                  30
Asp Val Ala Thr Ala Leu Pro Leu Gln Val Ala Pro Ser Ala Val Pro
        35                  40                  45
Met Asp Leu Arg Leu Asp His Gln Phe Ser Leu Pro Val Ala Glu Pro
    50                  55                  60
Ala Leu Arg Glu Gln Gln Leu Gln Gln Glu Leu Leu Ala Leu Lys Gln
65                  70                  75                  80
Lys Gln Gln Ile Gln Arg Gln Ile Leu Ile Ala Glu Phe Gln Arg Gln
                85                  90                  95
His Glu Gln Leu Ser Arg Gln His Glu Ala Gln Leu His Glu His Ile
            100                 105                 110
Lys Gln Gln Gln Glu Met Leu Ala Met Lys His Gln Gln Glu Leu Leu
```

```
        115                 120                 125
Glu His Gln Arg Lys Leu Glu Arg His Arg Gln Glu Gln Glu Leu Glu
    130                 135                 140
Lys Gln His Arg Glu Gln Lys Leu Gln Gln Leu Lys Asn Lys Glu Lys
145                 150                 155                 160
Gly Lys Glu Ser Ala Val Ala Ser Thr Glu Val Lys Met Lys Leu Gln
                165                 170                 175
Glu Phe Val Leu Asn Lys Lys Lys Ala Leu Ala His Arg Asn Leu Asn
            180                 185                 190
His Cys Ile Ser Ser Asp Pro Arg Tyr Trp Tyr Gly Lys Thr Gln His
        195                 200                 205
Ser Ser Leu Asp Gln Ser Ser Pro Pro Gln Ser Gly Val Ser Thr Ser
    210                 215                 220
Tyr Asn His Pro Val Leu Gly Met Tyr Asp Ala Lys Asp Asp Phe Pro
225                 230                 235                 240
Leu Arg Lys Thr Ala Ser Glu Pro Asn Leu Lys Leu Arg Ser Arg Leu
                245                 250                 255
Lys Gln Lys Val Ala Glu Arg Arg Ser Ser Pro Leu Leu Arg Arg Lys
            260                 265                 270
Asp Gly Pro Val Val Thr Ala Leu Lys Lys Arg Pro Leu Asp Val Thr
        275                 280                 285
Asp Ser Ala Cys Ser Ser Ala Pro Gly Ser Gly Pro Ser Ser Pro Asn
    290                 295                 300
Asn Ser Ser Gly Ser Val Ser Ala Glu Asn Gly Ile Ala Pro Ala Val
305                 310                 315                 320
Pro Ser Ile Pro Ala Glu Thr Ser Leu Ala His Arg Leu Val Ala Arg
                325                 330                 335
Glu Gly Ser Ala Ala Pro Leu Pro Leu Tyr Thr Ser Pro Ser Leu Pro
            340                 345                 350
Asn Ile Thr Leu Gly Leu Pro Ala Thr Gly Pro Ser Ala Gly Thr Ala
        355                 360                 365
Gly Gln Gln Asp Thr Glu Arg Leu Thr Leu Pro Ala Leu Gln Gln Arg
    370                 375                 380
Leu Ser Leu Phe Pro Gly Thr His Leu Thr Pro Tyr Leu Ser Thr Ser
385                 390                 395                 400
Pro Leu Glu Arg Asp Gly Gly Ala Ala His Ser Pro Leu Leu Gln His
                405                 410                 415
Met Val Leu Leu Glu Gln Pro Pro Ala Gln Ala Pro Leu Val Thr Gly
            420                 425                 430
Leu Gly Ala Leu Pro Leu His Ala Gln Ser Leu Val Gly Ala Asp Arg
        435                 440                 445
Val Ser Pro Ser Ile His Lys Leu Arg Gln His Arg Pro Leu Gly Arg
    450                 455                 460
Thr Gln Ser Ala Pro Leu Pro Gln Asn Ala Gln Ala Leu Gln His Leu
465                 470                 475                 480
Val Ile Gln Gln Gln His Gln Gln Phe Leu Glu Lys His Lys Gln Gln
                485                 490                 495
Phe Gln Gln Gln Gln Leu Gln Met Asn Lys Ile Ile Pro Lys Pro Ser
            500                 505                 510
Glu Pro Ala Arg Gln Pro Glu Ser His Pro Glu Glu Thr Glu Glu Glu
        515                 520                 525
Leu Arg Glu His Gln Ala Leu Leu Asp Glu Pro Tyr Leu Asp Arg Leu
    530                 535                 540
Pro Gly Gln Lys Glu Ala His Ala Gln Ala Gly Val Gln Val Lys Gln
545                 550                 555                 560
Glu Pro Ile Glu Ser Asp Glu Glu Glu Ala Glu Pro Pro Arg Glu Val
                565                 570                 575
Glu Pro Gly Gln Arg Gln Pro Ser Glu Gln Glu Leu Leu Phe Arg Gln
            580                 585                 590
Gln Ala Leu Leu Glu Gln Gln Arg Ile His Gln Leu Arg Asn Tyr
        595                 600                 605
Gln Ala Ser Met Glu Ala Ala Gly Ile Pro Val Ser Phe Gly Gly His
```

610 615 620
Arg Pro Leu Ser Arg Ala Gln Ser Ser Pro Ala Ser Ala Thr Phe Pro
625 630 635 640
Val Ser Val Gln Glu Pro Pro Thr Lys Pro Arg Phe Thr Thr Gly Leu
645 650 655
Val Tyr Asp Thr Leu Met Leu Lys His Gln Cys Thr Cys Gly Ser Ser
660 665 670
Ser Ser His Pro Glu His Ala Gly Arg Ile Gln Ser Ile Trp Ser Arg
675 680 685
Leu Gln Glu Thr Gly Leu Arg Gly Lys Cys Glu Cys Ile Arg Gly Arg
690 695 700
Lys Ala Thr Leu Glu Glu Leu Gln Thr Val His Ser Glu Ala His Thr
705 710 715 720
Leu Leu Tyr Gly Thr Asn Pro Leu Asn Arg Gln Lys Leu Asp Ser Lys
725 730 735
Lys Leu Leu Gly Ser Leu Ala Ser Val Phe Val Arg Leu Pro Cys Gly
740 745 750
Gly Val Gly Val Asp Ser Asp Thr Ile Trp Asn Glu Val His Ser Ala
755 760 765
Gly Ala Ala Arg Leu Ala Val Gly Cys Val Val Glu Leu Val Phe Lys
770 775 780
Val Ala Thr Gly Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro
785 790 795 800
Gly His His Ala Glu Glu Ser Thr Pro Met Gly Phe Cys Tyr Phe Asn
805 810 815
Ser Val Ala Val Ala Ala Lys Leu Leu Gln Gln Arg Leu Ser Val Ser
820 825 830
Lys Ile Leu Ile Val Asp Trp Asp Val His His Gly Asn Gly Thr Gln
835 840 845
Gln Ala Phe Tyr Ser Asp Pro Ser Val Leu Tyr Met Ser Leu His Arg
850 855 860
Tyr Asp Asp Gly Asn Phe Phe Pro Gly Ser Gly Ala Pro Asp Glu Val
865 870 875 880
Gly Thr Gly Pro Gly Val Gly Phe Asn Val Asn Met Ala Phe Thr Gly
885 890 895
Gly Leu Asp Pro Pro Met Gly Asp Ala Glu Tyr Leu Ala Ala Phe Arg
900 905 910
Thr Val Val Met Pro Ile Ala Ser Glu Phe Ala Pro Asp Val Val Leu
915 920 925
Val Ser Ser Gly Phe Asp Ala Val Glu Gly His Pro Thr Pro Leu Gly
930 935 940
Gly Tyr Asn Leu Ser Ala Arg Cys Phe Gly Tyr Leu Thr Lys Gln Leu
945 950 955 960
Met Gly Leu Ala Gly Gly Arg Ile Val Leu Ala Leu Glu Gly Gly His
965 970 975
Asp Leu Thr Ala Ile Cys Asp Ala Ser Glu Ala Cys Val Ser Ala Leu
980 985 990
Leu Gly Asn Glu Leu Asp Pro Leu Pro Glu Lys Val Leu Gln Gln Arg
995 1000 1005
Pro Asn Ala Asn Ala Val Arg Ser Met Glu Lys Val Met Glu Ile His
1010 1015 1020
Ser Lys Tyr Trp Arg Cys Leu Gln Arg Thr Thr Ser Thr Ala Gly Arg
1025 1030 1035 1040
Ser Leu Ile Glu Ala Gln Thr Cys Glu Asn Glu Glu Ala Glu Thr Val
1045 1050 1055
Thr Ala Met Ala Ser Leu Ser Val Gly Val Lys Pro Ala Glu Lys Arg
1060 1065 1070
Pro Asp Glu Glu Pro Met Glu Glu Glu Pro Pro Leu
1075 1080

<210> 13

<211> 3550
<212> DNA
<213> Homo sapiens

<400> 13
ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60
tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tggggtggct 120
ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtggatgtg 180
aagtcagaag ttcctgtggg cctggagccc atctcacctt tagacctaag gacagacctc 240
aggatgatga tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300
cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360
cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420
ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480
caagaacagg aagtagagag gcatcgcaga gaacagcagc ttcctcctct cagaggcaaa 540
gatagaggac gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600
ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660
cgccatccca agctctggta cacggctgcc caccacacat cattggatca aagctctcca 720
ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaag 780
gatgatttcc cccttcgaaa aactgcctct gagcccaact tgaaggtgcg gtccaggtta 840
aaacagaaag tggcagagag gagaagcagc cccttactca ggcggaagga tggaaatgtt 900
gtcacttcat tcaagaagcg aatgtttgag gtgacagaat cctcagtcag tagcagttct 960
ccaggctctg gtcccagttc accaaacaat gggccaactg gaagtgttac tgaaaatgag 1020
acttcggttt tgccccctac ccctcatgcc gagcaaatgg tttcacagca acgcattcta 1080
attcatgaag attccatgaa cctgctaagt ctttatacct ctccttcttt gcccaacatt 1140
accttggggc ttcccgcagt gccatcccag ctcaatgctt cgaattcact caaagaaaag 1200
cagaagtgtg agacgcagac gcttaggcaa ggtgttcctc tgcctgggca gtatggaggc 1260
agcatcccgg catcttccag ccaccctcat gttactttag agggaaagcc acccaacagc 1320
agccaccagg ctctcctgca gcatttatta ttgaaagaac aaatgcgaca gcaaaagctt 1380
cttgtagctg gtggagttcc cttacatcct cagtctcccт tggcaacaaa agagagaatt 1440
tcacctggca ttagaggtac ccacaaattg ccccgtcaca gacccctgaa ccgaacccag 1500
tctgcacctt tgcctcagag cacgttggct cagctggtca ttcaacagca acaccagcaa 1560
ttcttggaga agcagaagca ataccagcag cagatccaca tgaacaaact gctttcgaaa 1620
tctattgaac aactgaagca accaggcagt caccttgagg aagcagagga agagcttcag 1680
gggggaccagg cgatgcagga agacagagcg ccctctagtg gcaacagcac taggagcgac 1740
agcagtgctt gtgtggatga cacactggga caagttgggg ctgtgaaggt caaggaggaa 1800
ccagtggaca gtgatgaaga tgctcagatc caggaaatgg aatctgggga gcaggctgct 1860
tttatgcaac aggtaatagg caaagattta gctccaggat ttgtaattaa agtcattatc 1920
tgacctttcc tggaacccac gcacacacgt gcgctctctg tgcgccaagc tccgctggct 1980
gcggttggca tggatggatt agagaaacac cgtctcgtct ccaggactca ctcttcccct 2040
gctgcctctg ttttacctca cccagcaatg gaccgccccc tccagcctgg ctctgcaact 2100
ggaattgcct atgacccctt gatgctgaaa caccagtgcg tttgtggcaa ttccaccacc 2160
caccctgagc atgctggacg aatacagagt atctggtcac gactgcaaga aactgggctg 2220
ctaaataaat gtgagcgaat tcaaggtcga aaagccagcc tggaggaaat acagcttgtt 2280
cattctgaac atcactcact gttgtatggc accaaccccc tggacggaca gaagctggac 2340
cccaggatac tcctaggtga tgactctcaa aagttttttt cctcattacc ttgtggtgga 2400
cttggggtgg acagtgacac catttggaat gagctacact cgtccggtgc tgcacgcatg 2460
gctgttggct gtgtcatcga gctggcttcc aaagtggcct caggagagct gaagaatggg 2520
tttgctgttg tgaggccccc tggccatcac gctgaagaat ccacagccat ggggttctgc 2580
ttttttaatt cagttgcaat taccgccaaa tacttgagag accaactaaa tataagcaag 2640
atattgattg tagatctgga tgttcaccat ggaaacggta cccagcaggc cttttatgct 2700
gaccccagca tcctgtacat ttcactccat cgctatgatg aagggaactt tttccctggc 2760
agtggagccc caaatgaggt tcggtttatt tctttagagc cccactttta tttgtatctt 2820
tcaggtaatt gcattgcatg attacccctа attttcttgt cctttgctgg tgttttaaat 2880
tacacgagat tactgaattg tcccatggga ccaagaacca gtgcagaaca agtgcataac 2940
ccagagcact gtttgtcagg gaaggttggg ctgatttgat gtgttgtttg atgtttattt 3000
caagagctcc catgtgcttg ttttcctctc ttcttgcttt cttccatttg ctctcttctc 3060
tgcccaccgt ggtgtgtctt tctcttccca ggttggaaca ggccttggag aagggtacaa 3120
tataaatatt gcctggacag gtggccttga tcctcccatg ggagatgttg agtaccttga 3180
agcattcagg accatcgtga agcctgtggc caaagagttt gatccagaca tggtcttagt 3240
atctgctgga tttgatgcat tggaaggcca cacccctcct ctaggagggt acaaagtgac 3300
ggcaaaatgt tttggtcatt tgacgaagca attgatgaca ttggctgatg gacgtgtggt 3360
gttggctcta gaaggaggac atgatctcac agccatctgt gatgcatcag aagcctgtgt 3420

aaatgccctt ctaggaaatg agctggagcc acttgcagaa gatattctcc accaaagccc 3480
gaatatgaat gctgttattt ctttacagaa gatcattgaa attcaaagta tgtctttaaa 3540
gttctcttaa 3550

<210> 14
<211> 7699
<212> DNA
<213> Homo sapiens

<400> 14
cccattcgcc attcaggctg cgcaactgtt gggaagggcg atcggtgcgg gcctcttcgc 60
tattacgcca gctggcgaaa gggggatgtg ctgcaaggcg attaagttgg gtaacgccca 120
gggttttccc agtcacgacg ttgtaaaacg acggccagtg ccaagctgat ctaatcaata 180
ttggccatta gccatattat tcattggtta tatagcataa atcaatattg gctattggcc 240
attgcatacg ttgtatccat atcataatat gtacatttat attggctcat gtccaacatt 300
accgccatgt tgacattgat tattgactag ttattaatag taatcaatta cggggtcatt 360
agttcatagc ccatatatgg agttccgcgt tacataactt acggtaaatg gcccgcctgg 420
cgaccgccca gcgacccccg cccgttgacg tcaatagtga cgtatgttcc catagtaacg 480
ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg 540
gcagtacatc aagtgtatca tatgccaagt ccgcccccta ttgacgtcaa tgacggtaaa 600
tggcccgcct agcattatgc ccagtacatg accttacggg agtttcctac ttggcagtac 660
atctacgtat tagtcatcgc tattaccatg gtgatgcggt tttggcagta caccaatggg 720
cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga cgtcaatggg 780
agtttgtttt ggcaccaaaa tcaacgggac tttccaaaat gtcgtaataa ccccgccccg 840
ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag agctcgttta 900
gtgaaccgtc agaattcaag cttgcggccg cagatctatc gatctgcagg atatcaccat 960
gcacagtatg atcagctcag tggatgtgaa gtcagaagtt cctgtgggcc tggagcccat 1020
ctcaccttta gacctaagga cagacctcag gatgatgatg cccgtggtgg accctgttgt 1080
ccgtgagaag caattgcagc aggaattact tcttatccag cagcagcaac aaatccagaa 1140
gcagcttctg atagcagagt ttcagaaaca gcatgagaac ttgacacggc agcaccaggc 1200
tcagcttcag gagcatatca aggaacttct agccataaaa cagcaacaag aactcctaga 1260
aaaggagcag aaactggagc agcagaggca agaacaggaa gtagagaggc atcgcagaga 1320
acagcagctt cctcctctca gaggcaaaga tagaggacga gaaagggcag tggcaagtac 1380
agaagtaaag cagaagcttc aagagttcct actgagtaaa tcagcaacga aagacactcc 1440
aactaatgga aaaaatcatt ccgtgagccg ccatcccaag ctctggtaca cggctgccca 1500
ccacacatca ttggatcaaa gctctccacc ccttagtgga acatctccat cctacaagta 1560
cacattacca ggagcacaag atgcaaagga tgatttcccc cttcgaaaaa ctgcctctga 1620
gcccaacttg aaggtgcggt ccaggttaaa acagaaagtg gcagagagga gaagcagccc 1680
cttactcagg cggaaggatg gaaatgttgt cacttcattc aagaagcgaa tgtttgaggt 1740
gacagaatcc tcagtcagta gcagttctcc aggctctggt cccagttcac caaacaatgg 1800
gccaactgga agtgttactg aaaatgagac ttcggttttg ccccctaccc ctcatgccga 1860
gcaaatggtt tcacagcaac gcattctaat tcatgaagat tccatgaacc tgctaagtct 1920
ttatacctct ccttctttgc ccaacattac cttggggctt cccgcagtgc catcccagct 1980
caatgcttcg aattcactca aagaaaagca gaagtgtgag acgcagacgc ttaggcaagg 2040
tgttcctctg cctgggcagt atggaggcag catcccggca tcttccagcc accctcatgt 2100
tactttagag ggaaagccac ccaacagcag ccaccaggct ctcctgcagc atttattatt 2160
gaaagaacaa atgcgacagc aaaagcttct tgtagctggt ggagttccct tacatcctca 2220
gtctcccttg gcaacaaaag agagaatttc acctggcatt agaggtaccc acaaattgcc 2280
ccgtcacaga cccctgaacc gaacccagtc tgcacctttg cctcagagca cgttggctca 2340
gctggtcatt caacagcaac accagcaatt cttggagaag cagaagcaat accagcagca 2400
gatccacatg aacaaactgc tttcgaaatc tattgaacaa ctgaagcaac caggcagtca 2460
ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg atgcaggaag acagagcgcc 2520
ctctagtggc aacagcacta ggagcgacag cagtgcttgt gtggatgaca cactgggaca 2580
agttggggct gtgaaggtca aggaggaacc agtggacagt gatgaagatg ctcagatcca 2640
ggaaatggaa tctggggagc aggctgcttt tatgcaacag cctttcctgg aacccacgca 2700
cacacgtgcg ctctctgtgc gccaagctcc gctggctgcg gttggcatgg atggattaga 2760
gaaacaccgt ctcgtctcca ggactcactc ttcccctgct gcctctgttt tacctcaccc 2820
agcaatggac cgcccccctcc agcctggctc tgcaactgga attgcctatg acccccttgat 2880
gctgaaacac cagtgcgttt gtggcaattc caccacccac cctgagcatg ctggacgaat 2940
acagagtatc tggtcacgac tgcaagaaac tgggctgcta aataaatgtg agcgaattca 3000
aggtcgaaaa gccagcctgg aggaaataca gcttgttcat tctgaacatc actcactgtt 3060
gtatggcacc aaccccctgg acggacagaa gctggacccc aggatactcc taggtgatga 3120

ctctcaaaag ttttttttcct cattaccttg tggtggactt ggggtggaca gtgacaccat 3180
ttggaatgag ctacactcgt ccggtgctgc acgcatggct gttggctgtg tcatcgagct 3240
ggcttccaaa gtggcctcag gagagctgaa gaatgggttt gctgttgtga ggccccctgg 3300
ccatcacgct gaagaatcca cagccatggg gttctgcttt tttaattcag ttgcaattac 3360
cgccaaatac ttgagagacc aactaaatat aagcaagata ttgattgtag atctggatgt 3420
tcaccatgga aacggtaccc agcaggcctt ttatgctgac cccagcatcc tgtacatttc 3480
actccatcgc tatgatgaag ggaacttttt ccctggcagt ggagccccaa atgaggttgg 3540
aacaggcctt ggagaagggt acaatataaa tattgcctgg acaggtggcc ttgatcctcc 3600
catgggagat gttgagtacc ttgaagcatt caggaccatc gtgaagcctg tggccaaaga 3660
gtttgatcca gacatggtct tagtatctgc tggatttgat gcattggaag gccacacccc 3720
tcctctagga gggtacaaag tgacggcaaa atgttttggt catttgacga agcaattgat 3780
gacattggct gatggacgtg tggtgttggc tctagaagga ggacatgatc tcacagccat 3840
ctgtgatgca tcagaagcct gtgtaaatgc ccttctagga aatgagctgg agccacttgc 3900
agaagatatt ctccaccaaa gcccgaatat gaatgctgtt atttctttac agaagatcat 3960
tgaaattcaa agtatgtctt taaagttctc tggatccggt accagattac aaggacgacg 4020
atgacaagta gatcccgggt ggcatccctg tgacccctcc ccagtgcctc tcctggcctt 4080
ggaagttgcc actccagtgc ccaccagcct tgtcctaata aaattaagtt gcatcatttt 4140
gtctgactag gtgtcctcta taatattatg gggtggaggg gggtggtatg gagcaagggg 4200
cccaagttgg gaagacaacc tgtagggcct gcggggtcta ttcgggaacc aagctggagt 4260
gcagtggcac aatcttggct cactgcaatc tccgcctcct gggttcaagc gattctcctg 4320
cctcagcctc ccgagttgtt gggattccag gcatgcatga ccaggctcag ctaatttttg 4380
tttttttggt agagacgggg tttcaccata ttggccaggc tggtctccaa ctcctaatct 4440
caggtgatct acccaccttg gcctcccaaa ttgctgggat tacaggcgtg aaccactgct 4500
cccttccctg tccttctgat tttaaaataa ctataccagc aggaggacgt ccagacacag 4560
cataggctac ctgccatggc ccaaccggtg ggacatttga gttgcttgct tggcactgtc 4620
ctctcatgcg ttgggtccac tcagtagatg cctgttgaat tgggtacgcg gccagcttct 4680
gtggaatgtg tgtcagttag ggtgtggaaa gtccccaggc tccccagcag gcagaagtat 4740
gcaaagcatg catctcaatt agtcagcaac caggtgtgga aagtccccca ggctccccag 4800
caggcagaag tatgcaaagc atgcatctca attagtcagc aaccatagtc ccgcccctaa 4860
ctccgcccat cccgccccta actccgccca gttccgccca ttctccgccc catggctgac 4920
taattttttt tatttatgca gaggccgagg ccgcctcggc ctctgagcta ttccagaagt 4980
agtgaggagg cttttttgga ggcctaggct tttgcaaaaa gctcctcgag gaactgaaaa 5040
accagaaagt taattcccta tagtgagtcg tattaaattc gtaatcatgg tcatagctgt 5100
ttcctgtgtg aaattgttat ccgctcacaa ttccacacaa catacgagcc ggaagcataa 5160
agtgtaaagc ctggggtgcc taatgagtga gctaactcac attaattgcg ttgcgctcac 5220
tgcccgcttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg 5280
cggggagagg cggtttgcgt attgggcgct cttccgcttc ctcgctcact gactcgctgc 5340
gctcggtcgt tcggctgcgg cgagcggtat cagctcactc aaaggcggta atacggttat 5400
ccacagaatc aggggataac gcaggaaaga acatgtgagc aaaaggccag caaaaggcca 5460
ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc cctgacgagc 5520
atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta taaagatacc 5580
aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg 5640
gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcaatgc tcacgctgta 5700
ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg 5760
ttcagcccga ccgctgcgcc ttatccggta actatcgtct tgagtccaac ccggtaagac 5820
acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag 5880
gcggtgctac agagttcttg aagtggtggc ctaactacgg ctacactaga agaacagtat 5940
ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt agctcttgat 6000
ccggcaaaca aaccaccgct ggtagcggtg gtttttttgt ttgcaagcag cagattacgc 6060
gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacggggtct gacgctcagt 6120
ggaacgaaaa ctcacgttaa gggattttgg tcatgagatt atcaaaaagg atcttcacct 6180
agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat gagtaaactt 6240
ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtctatttc 6300
gttcatccat agttgcctga ctccccgtcg tgtagataac tacgatacgg gagggcttac 6360
catctggccc cagtgctgca atgataccgc gagacccacg ctcaccggct ccagatttat 6420
cagcaataaa ccagccagcc ggaagggccg agcgcagaag tggtcctgca actttatccg 6480
cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttaata 6540
gtttgcgcaa cgttgttgcc attgctacag gcatcgtggt gtcacgctcg tcgtttggta 6600
tggcttcatt cagctccggt tcccaacgat caaggcgagt tacatgatcc cccatgttgt 6660
gcaaaaaagc ggttagctcc ttcggtcctc cgatcgttgt cagaagtaag ttggccgcag 6720
tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg ccatccgtaa 6780
gatgcttttc tgtgactggt gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc 6840

gaccgagttg ctcttgcccg gcgtcaatac gggataatac cgcgccacat agcagaactt 6900
taaaagtgct catcattgga aaacgttctt cggggcgaaa actctcaagg atcttaccgc 6960
tgttgagatc cagttcgatg taacccactc gtgcacccaa ctgatcttca gcatctttta 7020
ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa 7080
taagggcgac acggaaatgt tgaatactca tactcttcct ttttcaatat tattgaagca 7140
tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag aaaaataaac 7200
aaataggggt tccgcgcaca tttccccgaa aagtgccacc tgacgcgccc tgtagcggcg 7260
cattaagcgc ggcgggtgtg gtggttacgc gcagcgtgac cgctacactt gccagcgccc 7320
tagcgcccgc tcctttcgct ttcttccctt cctttctcgc cacgttcgcc ggctttcccc 7380
gtcaagctct aaatcggggc atccctttag ggttccgatt tagtgcttta cggcacctcg 7440
accccaaaaa acttgattag ggtgatggtt cacgtagtgg gccatcgccc tgatagacgg 7500
tttttcgccc tttgacgttg gagtccacgt tctttaatag tggactcttg ttccaaactg 7560
gaacaacact caaccctatc tcggtctatt cttttgattt ataagggatt ttgccgattt 7620
cggcctattg gttaaaaaat gagctgattt aacaaaaatt taacgcgaat tttaacaaaa 7680
tattaaacgt ttacaattt 7699

<210> 15
<211> 7303
<212> DNA
<213> Homo sapiens

<400> 15
cccattcgcc attcaggctg cgcaactgtt gggaagggcg atcggtgcgg gcctcttcgc 60
tattacgcca gctggcgaaa gggggatgtg ctgcaaggcg attaagttgg gtaacgccca 120
gggttttccc agtcacgacg ttgtaaaacg acggccagtg ccaagctgat ctaatcaata 180
ttggccatta gccatattat tcattggtta tatagcataa atcaatattg gctattggcc 240
attgcatacg ttgtatccat atcataatat gtacatttat attggctcat gtccaacatt 300
accgccatgt tgacattgat tattgactag ttattaatag taatcaatta cggggtcatt 360
agttcatagc ccatatatgg agttccgcgt tacataactt acggtaaatg gcccgcctgg 420
cgaccgccca gcgacccccg cccgttgacg tcaatagtga cgtatgttcc catagtaacg 480
ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg 540
gcagtacatc aagtgtatca tatgccaagt ccgcccccta ttgacgtcaa tgacggtaaa 600
tggcccgcct agcattatgc ccagtacatg accttacggg agtttcctac ttggcagtac 660
atctacgtat tagtcatcgc tattaccatg gtgatgcggt tttggcagta caccaatggg 720
cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga cgtcaatggg 780
agtttgtttt ggcaccaaaa tcaacgggac tttccaaaat gtcgtaataa ccccgccccg 840
ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag agctcgttta 900
gtgaaccgtc agaattcaag cttgcggccg cagatctatc gatctgcagg atatcaccat 960
gcacagtatg atcagctcag tggatgtgaa gtcagaagtt cctgtgggcc tggagcccat 1020
ctcaccttta gacctaagga cagacctcag gatgatgatg cccgtggtgg accctgttgt 1080
ccgtgagaag caattgcagc aggaattact tcttatccag cagcagcaac aaatccagaa 1140
gcagcttctg atagcagagt ttcagaaaca gcatgagaac ttgacacggc agcaccaggc 1200
tcagcttcag gagcatatca aggaacttct agccataaaa cagcaacaag aactcctaga 1260
aaaggagcag aaactggagc agcagaggca agaacaggaa gtagagaggc atcgcagaga 1320
acagcagctt cctcctctca gaggcaaaga tagaggacga gaaagggcag tggcaagtac 1380
agaagtaaag cagaagcttc aagagttcct actgagtaaa tcagcaacga aagacactcc 1440
aactaatgga aaaaatcatt ccgtgagccg ccatcccaag ctctggtaca cggctgccca 1500
ccacacatca ttggatcaaa gctctccacc ccttagtgga acatctccat cctacaagta 1560
cacattacca ggagcacaag atgcaaagga tgatttcccc cttcgaaaaa ctgcctctga 1620
gcccaacttg aaggtgcggt ccaggttaaa acagaaagtg gcagagagga gaagcagccc 1680
cttactcagg cggaaggatg gaaatgttgt cacttcattc aagaagcgaa tgtttgaggt 1740
gacagaatcc tcagtcagta gcagttctcc aggctctggt cccagttcac caaacaatgg 1800
gccaactgga agtgttactg aaaatgagac ttcggttttg ccccctaccc ctcatgccga 1860
gcaaatggtt tcacagcaac gcattctaat tcatgaagat tccatgaacc tgctaagtct 1920
ttatacctct ccttctttgc ccaacattac cttggggctt cccgcagtgc catcccagct 1980
caatgcttcg aattcactca aagaaaagca gaagtgtgag acgcagacgc ttaggcaagg 2040
tgttcctctg cctgggcagt atggaggcag catcccggca tcttccagcc accctcatgt 2100
tactttagag ggaaagccac ccaacagcag ccaccaggct ctcctgcagc atttattatt 2160
gaaagaacaa atgcgacagc aaaagcttct tgtagctggt ggagttccct tacatcctca 2220
gtctcccttg gcaacaaaag agagaatttc acctggcatt agaggtaccc acaaattgcc 2280

ccgtcacaga cccctgaacc gaacccagtc tgcacctttg cctcagagca cgttggctca 2340
gctggtcatt caacagcaac accagcaatt cttggagaag cagaagcaat accagcagca 2400
gatccacatg aacaaactgc tttcgaaatc tattgaacaa ctgaagcaac caggcagtca 2460
ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg atgcaggaag acagagcgcc 2520
ctctagtggc aacagcacta ggagcgacag cagtgcttgt gtggatgaca cactgggaca 2580
agttggggct gtgaaggtca aggaggaacc agtggacagt gatgaagatg ctcagatcca 2640
ggaaatggaa tctggggagc aggctgcttt tatgcaacag cctttcctgg aacccacgca 2700
cacacgtgcg ctctctgtgc gccaagctcc gctggctgcg gttggcatgg atggattaga 2760
gaaacaccgt ctcgtctcca ggactcactc ttcccctgct gcctctgttt tacctcaccc 2820
agcaatggac cgcccccctcc agcctggctc tgcaactgga attgcctatg acccccttgat 2880
gctgaaacac cagtgcgttt gtggcaattc caccacccac cctgagcatg ctggacgaat 2940
acagagtatc tggtcacgac tgcaagaaac tgggctgcta aataaatgtg agcgaattca 3000
aggtcgaaaa gccagcctgg aggaaataca gcttgttcat tctgaacatc actcactgtt 3060
gtatggcacc aaccccctgg acggacagaa gctggacccc aggatactcc taggtgatga 3120
ctctcaaaag tttttttcct cattaccttg tggtggactt ggggtggaca gtgacaccat 3180
ttggaatgag ctacactcgt ccggtgctgc acgcatggct gttggctgtg tcatcgagct 3240
ggcttccaaa gtggcctcag gagagctgaa gaatgggttt gctgttgtga ggccccctgg 3300
ccatcacgct gaagaatcca cagccatggg gttctgcttt tttaattcag ttgcaattac 3360
cgccaaatac ttgagagacc aactaaatat aagcaagata ttgattgtag atctggatgt 3420
tcaccatgga aacggtaccc agcaggcctt ttatgctgac cccagcatcc tgtacatttc 3480
actccatcgc tatgatgaag ggaacttttt ccctggcagt ggagccccaa atgaggttcg 3540
gtttatttct ttagagcccc acttttattt gtatctttca ggtaattgca ttgcaggatc 3600
cggtaccaga ttacaaggac gacgatgaca agtagatccc gggtggcatc cctgtgaccc 3660
ctccccagtg cctctcctgg ccttggaagt tgccactcca gtgcccacca gccttgtcct 3720
aataaaatta agttgcatca ttttgtctga ctaggtgtcc tctataatat tatggggtgg 3780
aggggggtgg tatggagcaa ggggcccaag ttgggaagac aacctgtagg gcctgcgggg 3840
tctattcggg aaccaagctg gagtgcagtg gcacaatctt ggctcactgc aatctccgcc 3900
tcctgggttc aagcgattct cctgcctcag cctcccgagt tgttgggatt ccaggcatgc 3960
atgaccaggc tcagctaatt tttgtttttt tggtagagac ggggtttcac catattggcc 4020
aggctggtct ccaactccta atctcaggtg atctacccac cttggcctcc caaattgctg 4080
ggattacagg cgtgaaccac tgctcccttc cctgtccttc tgattttaaa ataactatac 4140
cagcaggagg acgtccagac acagcatagg ctacctgcca tggcccaacc ggtgggacat 4200
ttgagttgct tgcttggcac tgtcctctca tgcgttgggt ccactcagta gatgcctgtt 4260
gaattgggta cgcggccagc ttctgtggaa tgtgtgtcag ttagggtgtg gaaagtcccc 4320
aggctcccca gcaggcagaa gtatgcaaag catgcatctc aattagtcag caaccaggtg 4380
tggaaagtc cccaggctcc ccagcaggca gaagtatgca aagcatgcat ctcaattagt 4440
cagcaaccat agtcccgccc ctaactccgc ccatcccgcc cctaactccg cccagttccg 4500
cccattctcc gccccatggc tgactaattt ttttatttta tgcagaggcc gaggccgcct 4560
cggcctctga gctattccag aagtagtgag gaggcttttt tggaggccta ggcttttgca 4620
aaaagctcct cgaggaactg aaaaaccaga aagttaattc cctatagtga gtcgtattaa 4680
attcgtaatc atggtcatag ctgtttcctg tgtgaaattg ttatccgctc acaattccac 4740
acaacatacg agccggaagc ataaagtgta aagcctgggg tgcctaatga gtgagctaac 4800
tcacattaat tgcgttgcgc tcactgcccg ctttccagtc gggaaacctg tcgtgccagc 4860
tgcattaatg aatcggccaa cgcgcgggga gaggcggttt gcgtattggg cgctcttccg 4920
cttcctcgct cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc 4980
actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt 5040
gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc 5100
ataggctccg cccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa 5160
acccgacagg actataaaga taccaggcgt ttccccctgg aagctccctc gtgcgctctc 5220
ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg 5280
cgctttctca atgctcacgc tgtaggtatc tcagttcggt gtaggtcgtt cgctccaagc 5340
tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc 5400
gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca 5460
ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact 5520
acggctacac tagaagaaca gtatttggta tctgcgctct gctgaagcca gttaccttcg 5580
gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt 5640
ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct 5700
tttctacggg gtctgacgct cagtggaacg aaaactcacg ttaagggatt ttggtcatga 5760
gattatcaaa aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttaaatcaa 5820
tctaaagtat atatgagtaa acttggtctg acagttacca atgcttaatc agtgaggcac 5880
ctatctcagc gatctgtcta tttcgttcat ccatagttgc ctgactcccc gtcgtgtaga 5940
taactacgat acgggagggc ttaccatctg gccccagtgc tgcaatgata ccgcgagacc 6000

```
cacgctcacc ggctccagat ttatcagcaa taaaccagcc agccggaagg gccgagcgca 6060
gaagtggtcc tgcaacttta tccgcctcca tccagtctat taattgttgc cgggaagcta 6120
gagtaagtag ttcgccagtt aatagtttgc gcaacgttgt tgccattgct acaggcatcg 6180
tggtgtcacg ctcgtcgttt ggtatggctt cattcagctc cggttcccaa cgatcaaggc 6240
gagttacatg atcccccatg ttgtgcaaaa aagcggttag ctccttcggt cctccgatcg 6300
ttgtcagaag taagttggcc gcagtgttat cactcatggt tatggcagca ctgcataatt 6360
ctcttactgt catgccatcc gtaagatgct tttctgtgac tggtgagtac tcaaccaagt 6420
cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgtca atacgggata 6480
ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt tcttcggggc 6540
gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaaccc actcgtgcac 6600
ccaactgatc ttcagcatct tttactttca ccagcgtttc tgggtgagca aaaacaggaa 6660
ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa atgttgaata ctcatactct 6720
tcctttttca atattattga agcatttatc agggttattg tctcatgagc ggatacatat 6780
ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc cgaaaagtgc 6840
cacctgacgc gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt acgcgcagcg 6900
tgaccgctac acttgccagc gccctagcgc ccgctccttt cgctttcttc ccttcctttc 6960
tcgccacgtt cgccggcttt ccccgtcaag ctctaaatcg ggcatccct ttagggttcc 7020
gatttagtgc tttacggcac ctcgacccca aaaaacttga ttagggtgat ggttcacgta 7080
gtgggccatc gccctgatag acggtttttc gccctttgac gttggagtcc acgttcttta 7140
atagtggact cttgttccaa actggaacaa cactcaaccc tatctcggtc tattcttttg 7200
atttataagg gattttgccg atttcggcct attggttaaa aaatgagctg atttaacaaa 7260
aatttaacgc gaattttaac aaaatattaa acgtttacaa ttt                   7303
```

<210> 16
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer used to amplify human DNA

<400> 16
ccatggaaac ggtacccagc aggc 24

<210> 17
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer used to amplify human DNA

<400> 17
cactccatcg ctatgatgaa ggg 23

<210> 18
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer used to amplify human DNA

<400> 18
agttcccttc atcatagcga tgg 23

<210> 19
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer used to amplify human DNA

<400> 19
aatgtacagg atgctggggt 20

<210> 20
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer used to amplify human DNA

<400> 20
cccttgtagc tggtggagtt ccctt 25

<210> 21
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer used to amplify human DNA

<400> 21
tgtgtcatcg agctggcttc 20

<210> 22
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer used to amplify human DNA

<400> 22
atcttctgca agtggctcca 20

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 December 2002 (27.12.2002)

PCT

(10) International Publication Number
WO 02/102984 A3

(51) International Patent Classification[7]: C12N 9/78, 9/00, 9/14, 1/20, 15/00, C07H 21/04

(21) International Application Number: PCT/US02/19051

(22) International Filing Date: 14 June 2002 (14.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/298,173	14 June 2001 (14.06.2001)	US
60/311,686	10 August 2001 (10.08.2001)	US
60/316,995	4 September 2001 (04.09.2001)	US

(71) Applicant *(for all designated States except US)*: **SLOAN-KETTERING INSTITUTE FOR CANCER RESEARCH** [US/US]; 1275 York Avenue, New York, NY 10021 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **RICHON, Victoria** [US/US]; 160 Theodore Fremd Street, #A11, Rye, NY 10580 (US). **ZHOU, Xianbo** [CN/US]; 43 Bradley Street, Dobbs Ferry, NY 10522 (US). **RIFKIND, Richard, A.** [US/US]; 425 East 58th Street, #48A, New York, NY 10022 (US). **MARKS, Paul, A.** [US/US]; 7 Rossiter Road, Washington, CT 06793 (US).

(74) Agents: **BROOK, David, E.** et al.; Hamilton, Brook, Smith & Reynolds, P.C., 530 Virginia Road, P.O. Box 9133, Concord, MA 01742-9133 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report: 13 November 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/102984 A3

(54) Title: HDAC9 POLYPEPTIDES AND POLYNUCLEOTIDES AND USES THEREOF

(57) Abstract: The present invention features substantially pure HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), an HDRP(ΔNLS) polypeptides, and isolated nucleic acid molecules encoding those polypeptides. The present invention also features vectors containing*HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS),* and *HDRP(ΔNLS)* nucleic acid sequences, and cells containing those vectors.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/19051

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 9/78, 9/00, 9/14, 1/20, 15/00; C07H 21/04
US CL : 435/227, 183, 195, 252.3, 320.1; 536/23.2

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/227, 183, 195, 252.3, 320.1; 536/23.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
STN AND WEST. Sequence search in Swissprot, EST, N-GeneSeq, PIR_71, SPTREMBL & issued US patents.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NAGASE et al. Prediction of Coding Sequences of Unidentified Human Genes. XI. The Complete Sequences of 100 New cDNA Clones from Brain Which Code for Large Proteins in Vitro. DNA Research November 1998, Vol 5, pages 277-286. See Table 1, Accession No. AB018287 is 58.8% similar to DNA sequence of SEQ IF SEQ ID NO : 1, claim 4 (g).	4
A, P	ZHOU et al. Cloning and Characterization of a histone deacetylase, HDAC9. PNAS, 11 September 2001, Vol. 98, No. 19, pages 10572-10577.	1-9, 29
A	WANG et al. HDAC4, a Human Histone Deacetylase Related to Yeast HDA1, Is a Transcriptional Corepressor. Molecular and Cellular Biology, November 1999, Vol. 19, No. 11, pages 7816-7827.	1-9, 29

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 October 2002 (30.10.2002)	13 MAR 2003
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer T.K. Saidha Tekchand Saidha Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/19051

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-9 & 29 (SEQ ID NOS : 1 & 2)

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1-9, 29, drawn to isolated nucleic acid, the encoded protein and protein composition.

Group II, claim(s) 10, drawn to antibody.

Group III, claim(s) 11-13, drawn to a method of identifying a compound - modulate DNA expression.

Group IV, claim(s) 14-19, 33, drawn to a method of identifying a compound that modulate enzymatic activity.

Group V, claim(s) 20-25, 34, drawn to a method of identifying a compound that modulate transcriptional repression activity of the polypeptide.

Group VI, claim(s) 26-27, drawn to a method of identifying a compound that modulate expression of a nucleic acid molecule.

Group VII, claim(s) 28, drawn to a method of identifying a polypeptide that interacts with a polypeptide of claim 1 in a two-hybrid system.

Group VIII, claim(s) 30-32, drawn to a method of diagnosing a cell proliferation disease.

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack unity of invention because they are not so linked as to form a single general inventive concept under PCT Rule 13.1.

In order for more than one species to be examined, the appropriate additional examination fees must be paid. The species are as follows:

1. SEQ ID NO : 1 and 2 [HDAC9].
2. SEQ ID NO : 3 and 4 [HDAC9a].
3. SEQ ID NO : 5 and 6 [HDAC9-ΔNLS].
4. SEQ ID NO : 7 and 8 [HDAC9a-ΔNLS].
5. SEQ ID NO : 9 and 10 [HDRP-ΔNLS].

The claims are deemed to correspond to the species listed above in the following manner:

Each of the claims listed in groups I-VIII correspond to each of the 5 species which are structurally distinct.

The following claim(s) are generic: 1-5.

The inventions listed as Groups I-VIII do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Group I has a special technical feature of the nucleotide sequence encoding a specific histone deacetylase which Groups II-VIII do not share; Group II has a special technical feature of the antibody to a specific histone deacetylase which Groups I & III-VIII do not share; Groups III-VIII employ nucleic acid or polypeptide in various method of identifying compounds or polypeptides for distinct uses. Further, in view of 37 CFR 1.475 (b), when claims corresponding to different categories of inventions are present then only (3) applies and additional methods of use are deemed to lack unity.

The species listed above do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: The various species correspond to nucleic acid and polypeptide sequences which are structurally and in activity distinct from each other, therefore lack the same or corresponding special technical feature.